UNCERTAINTY IN HYDROLOGICAL MODELLING: A CASE STUDY IN THE TERN CATCHMENT, SHROPSHIRE, UK

Thesis submitted for the degree of
Doctor of Philosophy at UCL

REBEKAH ELIZABETH LYN ROCHESTER

Department of Geography
UCL

May 2010

I, Rebekah Elizabeth Lyn Rochester confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis explores a range of uncertainty issues within the commonly used hydrological modelling framework. It assesses the extent that choices made during model construction and calibration result in different model outputs and aims to assess whether it is possible to develop a modelling protocol better than the rest.

Using the 876.36 km^2 Tern catchment, Shropshire, UK, and the physically-based, distributed modelling code, MIKE SHE, the research draws on large volumes of secondary data and provides a comprehensive catchment review and conceptual model. Two hydrological models of differing spatial complexities are developed and subject to different parameterisations, sensitivity analyses, and calibration methods (manual and automatic). Results are assessed at different locations within the catchment.

Six models developed with different protocols result in minimal intra-model uncertainty. Nash-Sutcliffe NSE varies between 0.69–0.79 for discharge at the catchment outlet. Differences between spatial representations are more apparent at internal gauging stations; despite this similar performing models are developed for both spatial representations. Multi-objective automatic calibration produces models which provide more balanced representation of observed data as shown by results of validation. However, it is not possible to statistically identify any of the modelling protocols as better than the rest. Results suggest the amount a particular statistic is used within the calibration will influence other performance statistics. Therefore an independent summary score measure is also developed to assess performance.

Intra-model uncertainty is assessed for the six models for eight UKCIP02 climate change scenarios. Results suggest increases in intra-model uncertainty at a similar magnitude as potential impacts of climate change. The research suggests careful choices about the modelling protocol need to be addressed at the outset of any hydrological modelling, with attention given to the uncertainties that may result of decisions made by the modeller – especially if using models in impact studies.

Acknowledgements

People that know me will know how much of a long road this PhD has been. Thank you all for supporting me on the way. I am grateful for the three years of funding from the Geography department at UCL. It is no surprise to some that I enjoyed this computer support work, I'm sure my experience of it will last a lifetime!

I want to acknowledge the people and organisations that have given advice or provided the data that have been used in the thesis. Tabitha Sudworth (CEH Wallingford) for organising the licences and providing links for much of the LOCAR catchment data. I am also grateful to Gary Fulcher, Claire French and Amy Burt (Environment Agency), Douglas Graham (DHI-WE, Denmark), Sara Larman at Cranfield University, Andrew Mackenzie (British Geological Survey), John Catt (UCL) and John Hollis for an interesting meeting about soil. I am also grateful for the free use of the UKCIP02 climate change data, and the British Atmospheric Data Centre (BADC).

Thanks also to my friends and family, in and outside work, for making the process that little bit lighter and brighter: Raj, Lucinda, Adam, Hong, Jess, Clarkee, Kathryn, Grandma Beverley as well as all the general office girls, to name just a few. Thanks also to everyone at TEP for being patient and supportive. Thanks Miles for your help, and Helene Burningham for your enquiries about my work and understanding of my writing block.

I also want to express my thanks to my supervisor, Julian Thompson. I have often sighed after supervisory meetings but I do appreciate the support you have given me, and also for believing in me all those years ago as an undergraduate! Julian also quite rightly pointed out that I should acknowledge Birdy, Shuilong and Takea. Without them the work would have been impossible, even if I often wanted to throw them out the window. Thanks also to Mike Acreman, my second supervisor, your comments and suggestions on the wider approach have been useful.

Ultimately this thesis exists thanks to Nick, Sarah, my Mum and Dad, and my soon to be in-laws. I honestly can never express my gratitude for your patience, or your energies when mine failed, or thank you all enough for your support. I know I've made life difficult at times – and in the context of a whole thesis, these words seem so little, but they are the most important of the whole document. Thank you.

Table of contents

Glossary of terms used in the thesis

Automatic calibration at basin outlet – One of two automatic calibration methods used in this thesis. The calibration is carried out within Autocal, a component of the MIKE ZERO framework. Using the Shuffled Complex Evolution method, the optimisation of the parameters are assessed using the performance of the RMSE statistic of river flow at the basin outlet gauging station, Walcot. This automatic calibration method is undertaken for both the homogenous and distributed models.

Automatic calibration at multi-location and criteria – The second of two automatic calibration methods used in this thesis. The calibration is carried out within Autocal, a component of the MIKE ZERO framework. Using the Shuffled Complex Evolution method, the optimisation of the parameters are assessed using an objective function that equally weights the performance of the RMSE statistic of river flow at four gauging stations and seven groundwater level boreholes. This automatic calibration method is undertaken for both the homogenous and distributed models.

Automatic sensitivity analysis – A procedure undertaken within the Autocal component of MIKE ZERO for the distributed model in which the values of 60 model parameters are individually perturbed and the resulting RMSE of the simulation compared to an initial control simulation in which none of the parameters are perturbed.

Data uncertainty – Although not directly assessed in the thesis, reference is made to this form of uncertainty which is defined as potentially suspect or erroneous data (as well as general inherent uncertainty associated with any data collection) and subsequently input into the hydrological models.

Distributed model – The second of two spatial complexities of model types documented in this thesis and described in Chapter 4. Within this type of model many of the variables and parameters (although not all) vary spatially at the grid cell scale of 1km × 1km.

Expert Elicitation – A term expressed by Refsgaard et al (2006) and developed in this thesis. It refers to a subjective but specialist evaluation of model performance by the

hydrological modeller. The expert elicitation is undertaken by assessment of simulated hydrographs and groundwater levels and includes evaluation of key components such as peak flows, base flows and timing of hydrological events compared to the observed data. The summary score is a quantitative measure that has been developed in association with expert elicitation.

Grid cell – The division used within the models to define one element of the model to which different parameter values can be attributed. The grid cell scale used in the models is 1km × 1km.

Homogeneous model – One of the two spatial complexities of model types documented in this thesis and described in Chapter 4. The model is not lumped, as it still includes 1km gridded topography data, and is coupled to a 1D hydraulic MIKE 11 model. However, the majority of other input data use the catchment average value in each 1km × 1km grid cell.

Indirect acquisition of parameter values – The process of deriving parameter values or extents from pre-published literature. The values are often representative of typical values rather than specific to the catchment. For example, specific yield measurements for different media published by Johnson (1967).

Initial manual calibration – One of three calibration methods documented in the thesis. This calibration method required supervised variation of each separate model parameter, until representative values were defined and found to result in the model with the best quantitative and qualitative performance possible.

Manual sensitivity analysis – A procedure undertaken for the homogenous model in Chapter 5 where the values of 12 model parameters are subject to individual percentage perturbations to assess the extent to which each parameter influences the simulation of river flow and groundwater levels.

Model Equifinality – A term describing the recognition that there may be more than one modelling protocol in which resulting model performance is quantitatively and qualitatively similar.

Modelling Protocol – Defined in this thesis as the methodological framework chosen by the hydrological modeller to represent the catchment and calibrate the model. These choices include the way in which variables are spatially represented, the adopted grid size used in the model, the method used to calibrate and test the models and the choice of performance measures used to assess the models.

Optimal models – Are the best calibrated model from each modelling protocol assessed in this thesis. For the automatically calibrated models, the optimal model is the statistically best model derived from the optimisation method. In total there are six optimum models developed in this thesis.

Optimisation – The procedure of improving the performance of the models during the calibration process to the best level achievable. The optimisation is iterative and governed by a pre-defined criteria (e.g. improving the RMSE to the lowest score).

Parameter – The second type of two inputs into the hydrological models. Defined as attribute data that often require calibration or are acquired indirectly from literature defined typical values. For example, the four types of required soil hydraulic information that describe each soil class in the models.

Parameter Equifinality – Recognition that there may be more than one set of parameter values within the parameter space that result in models which are quantitatively and qualitatively similar, all of which can be described as calibrated.

Parameter space - the hypothetical region defined by the lower and upper values of all optimisable parameters.

Parameter uncertainty - A type of uncertainty in hydrological modelling due to a result of defining parameter values that are not representative of the attributes or processes they seek to describe. Parameter uncertainty may be due to scaling effects (the value not being representative at the grid cell scale), or may also be due to data uncertainty.

Qualitative performance – Includes all descriptive assessment of model performance. This is undertaken in detail in chapters 5 and 6 where any results of daily, monthly and annual river flow or groundwater levels are analysed in comparison to observed data.

Quantitative performance – Includes all numerical assessment of the model performance. This may include specialist statistics such as the Nash-Sutcliffe R2 or simple statistics such as mean river flow or groundwater levels. The summary score measure that is developed in the thesis is a means of assessing the quantitative performance of qualitative indicators.

Summary score – A method of model performance developed within the thesis. The score provides a means of quantifying the qualitative performance of the models by means of 'expert elicitation'.

Variable – One of two types of input into the models. Defined as 'input data' such as precipitation, evapotranspiration, or the geographical distribution of data such as for solid geology or topography. Variables are not subject to calibration.

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Introduction

Water management is increasingly recognising the need for integrated approaches to problem solving (Thompson and Hollis, 1997; United Nations, 1992). Hydrological models can provide a framework for this approach as they have the ability to couple and include all components of the hydrological system. They are also beneficial as they can be used to examine a range of scenarios whether that is, for example, to assess the impacts of climate change or alternative water resources management issues. As a result, hydrological modelling research and application are being driven by the increasing pressure being placed on global water resources as a result of rapid growth in population, economy and climate change.

The stress on water resources is exacerbated by a lack of co-ordinated management and governance (UNESCO, 2003) and there are numerous issues that urgently need attention ranging from water pollution and flooding, to problems of water scarcity and drought. Catchment hydrological models are therefore frequently being employed when seeking to address these global problems and issues.

The growth in research and application of hydrological models in recent years has therefore been primarily application driven, with the end users pushing for answers to

problems. Emphasis has not always been on careful and structured development that seeks to minimise uncertainty and error within the modelling protocol outlined below (Silberstein, 2006).

As defined by Klemes (1986, p.14) 'a hydrological simulation model is a mathematical model aimed at synthesising a (continuous) record of some hydrological variable Y, for a period T, from available and concurrent records of other variables X, Z..'. The concept of a hydrological model is considered to be based on the foundations of the hydrological system as a simplification of reality, and allows the general flows and stores of water to be expressed diagrammatically.

Figure 1.1 provides a summary of the protocol usually followed when undertaking hydrological modelling. After the identification of the research question, the first requirement in the modelling process is model conceptualisation (Lee, 1993), where the modeller is required to understand the basic hydrology of the system or catchment, review the available data and to decide on the basic structure of the model. This is often undertaken by consideration of the processes and stores associated with the land phase of the cycle in space and time. The hydrological model is usually represented by parameters (the parameterisation) that together make up each process or store within the model. Following model conceptualisation, a suitable model code must then be chosen that best suits the research problem and the available data (Refsgaard, 1997; 2000).

The process of modelling subsequently employs the ability of computers to manipulate large volumes of input data, and is based on careful logical programming (Shaw, 1994). Once the hydrological model has been constructed, performance criteria are defined in order to assess model performance. These criteria should once again best reflect the purpose of the research. For example, if the objective is to model flood discharges then suitable statistics that are concerned with error of peak flow simulation should be employed. In contrast, if the aim is to accurately simulate river flows over the full range of observed conditions then alternative performance criteria may be selected (Refsgaard, 2000). Both quantitative and qualitative performance criteria can be used including specific river flow test statistics such as the widely used Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970).

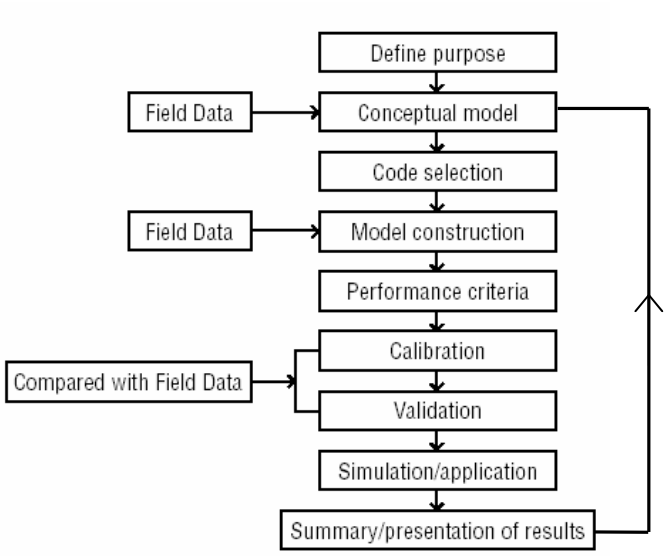


Figure 1.1. The steps in hydrological modelling protocol
(modified from Refsgaard, 1997)

The following stage is to calibrate the model which involves the refinement of parameters to values that are best representative of catchment characteristics, whilst at the same time seeking to maximise the ability of the simulation with regard to the defined performance criteria. Calibration methods can be manual (e.g. Refsgaard, 1997) or automatic using optimisation algorithms (Madsen, 2003). Once the model is calibrated, a testing or validation stage is necessary to assess model performance with an independent observed set of data outside of that used in the calibration. For example, testing is typically undertaken by the 'split sample' (Klemes, 1986) approach of using a separate time period and then re-assessing the performance criteria for that period. Additionally this phase can be undertaken with additional observed internal site data within the catchment not originally used in the calibration stage (such as river flow, groundwater levels, soil moisture or snow depth).

The initial stages of conceptualisation, model development and testing are critical within hydrological modelling and can often require a large proportion of the total project time taken. As shown in Figure 1.1, the modelling protocol indicates that only in the last two steps are hydrological models used for simulation and application purposes (Refsgaard, 1997). It is important to choose the correct modelling code for the problem and represent the catchment in a way that will result in the best solution for the initial objectives. Figure 1.1 shows that the calibrated hydrological model is then frequently used for simulating alternative scenarios in order to address the objectives of the

research. For example, in cases where the aim is to assess potential impacts of climate change on catchment water resources, perturbed input climate data of a particular future scenario are used to drive the model, and the output measure of river flow/groundwater level then compared to the original model output (Arnell and Reynard, 1996; Thodsen, 2007; Mernild et al., 2008; Christensen et al., 2004; Thompson et al., 2009). Estimated changes can then be derived and passed on to decision makers that seek to manage the impact of climate change. Other scenario testing that is also undertaken includes the modification to internal model structure, for example with scenarios of different groundwater abstraction (e.g. Shepley et al., 2009).

1.2. Aims and objectives of the research

As a result of the need for problem solving driving the development of hydrological models; many different model codes have been constructed with alternative spatial and process complexities in order to address the needs of specific research objectives. Added to the need for problem solving, continued quests to formalise knowledge and advance scientific understanding of processes within the hydrological system has resulted in the development of an increasing number of refined and more complex models. This has also been facilitated by the rapid development of computing power (Singh et al., 1995). As the number of models in operation increases however, the choice of which hydrological model code to use becomes greater and there appears to be comparatively few studies which compare model codes, methods of model set-up, and types of uncertainty in comparison to the volume of application driven modelling. As a result of this identified problem, the main aim of this thesis is:

To assess using an ensemble approach, the extent to which different choices in the construction and calibration process (key aspects of the modelling protocol) result in differences in simulated model outputs, and consequently whether it is possible to select a best modelling protocol from those assessed.

The research is undertaken using the 876.36 km^2 Tern Catchment in Shropshire, UK, as a case study. The key questions below have been highlighted in the literature as fundamental issues of uncertainty within the hydrological modelling protocol, and are used in this thesis in order to achieve the main aim of the research:

1. What effects do different spatial representations have on model outputs?

This is undertaken by the construction of two catchment scale hydrological models for the same catchment using the same model code, but with different levels of spatial descretisation. River flow and groundwater levels simulated by the models, and the model performance statistics are compared at a number of locations, both at the basin outlet and internally within the catchment.

2. How do different methods of model calibration affect the performance of river flows and groundwater levels?

For both models of different spatial complexities, a range of different methods of model calibration (manual and two automatic methods) are employed. In order to ensure the results can be compared, the two models of different spatial complexities are, as far as possible, subject to the same calibration methods. The performances of different methods of calibration are assessed using the same criteria – for river flow and groundwater levels at a number of sites internally, and at the basin outlet in the case of river flow. As a result of the different calibration methods six optimally calibrated models for the catchment are established (three calibrations for each spatial representation).

3. How different are river flows and groundwater levels as a result of parameter equifinality?

For the models that are subject to automatic calibration, the equifinality of the parameter space is investigated by the selection of a suite of calibrated models, each with different parameter values. The objective is to recognise that parameter equifinality exists and is an important uncertainty within hydrological model development.

4. To what extent do different performance statistics suggest different abilities of models?

Routinely used quantitative performance statistics such as Nash-Sutcliffe (NSE) (Nash and Sutcliffe, 1970), Root Mean Square Error (RMSE) and Correlation coefficient (R), are compared. An assessment is undertaken to assess whether results suggest that any one of the modelling protocols is better than the others. A further statistic, the summary

score, a measure which attempts to score more qualitative aspects of simulated river flow hydrographs and groundwater levels is also included in the assessment.

5. How do measures of performance internally within the catchment, as well river flow at the basin outlet suggest different abilities of models?

Assessment of model performance is undertaken for both river flows and groundwater levels at a number of locations, therefore using the multi-objective framework that is suggested for distributed models (Refsgaard, 1997; 2000). Assessment of river flow will be assessed at the basin outlet as well as major tributaries and minor upstream areas of the catchment. Groundwater levels are assessed at a range of locations across the catchment, in areas of differing geologies. The inclusion of a range of model performance criteria and their application to both river flow and groundwater levels facilitates a more comprehensive assessment of overall model performance.

A secondary aim of the thesis is to use the results to explore the range of intra-model uncertainty when using the models for the common application of climate change impact assessment on water resources. As already noted, a large number of climate change impact assessments compile uncertainty from a range of future scenarios or use data projected from a range of different climate models. This work uses an ensemble approach based on a range of calibrated models to compare intra-hydrological model uncertainty with the projected impact of climate change in the catchment. The specific research questions that are addressed in this part of the thesis are:

6. What are the projected impacts of climate change simulated for the Tern catchment, using UKCIP02 data, for both river flows and groundwater levels at a range of locations?

7. What is the uncertainty and range of simulated river flows and groundwater levels between the six calibrated hydrological models for different climate change scenarios and two time-slices (2050s and 2080s)?

8. What is the magnitude of the intra-model uncertainty compared to the projected impacts of climate change, and how important is intra-model variability when using the models to simulate the potential effects of climate change?

1.3. Research strategy

Given the research questions above that are used to assess the thesis aim, Figure 1.2 outlines a conceptual strategy that the research follows. This has been developed from the routine modelling protocol that was shown in Figure 1.1.

The conceptual strategy divides the research into nine individual components which form the basis of each chapter within the thesis. This conceptual strategy is described in detail in Section 1.5 that gives an overview of the structure of the thesis.

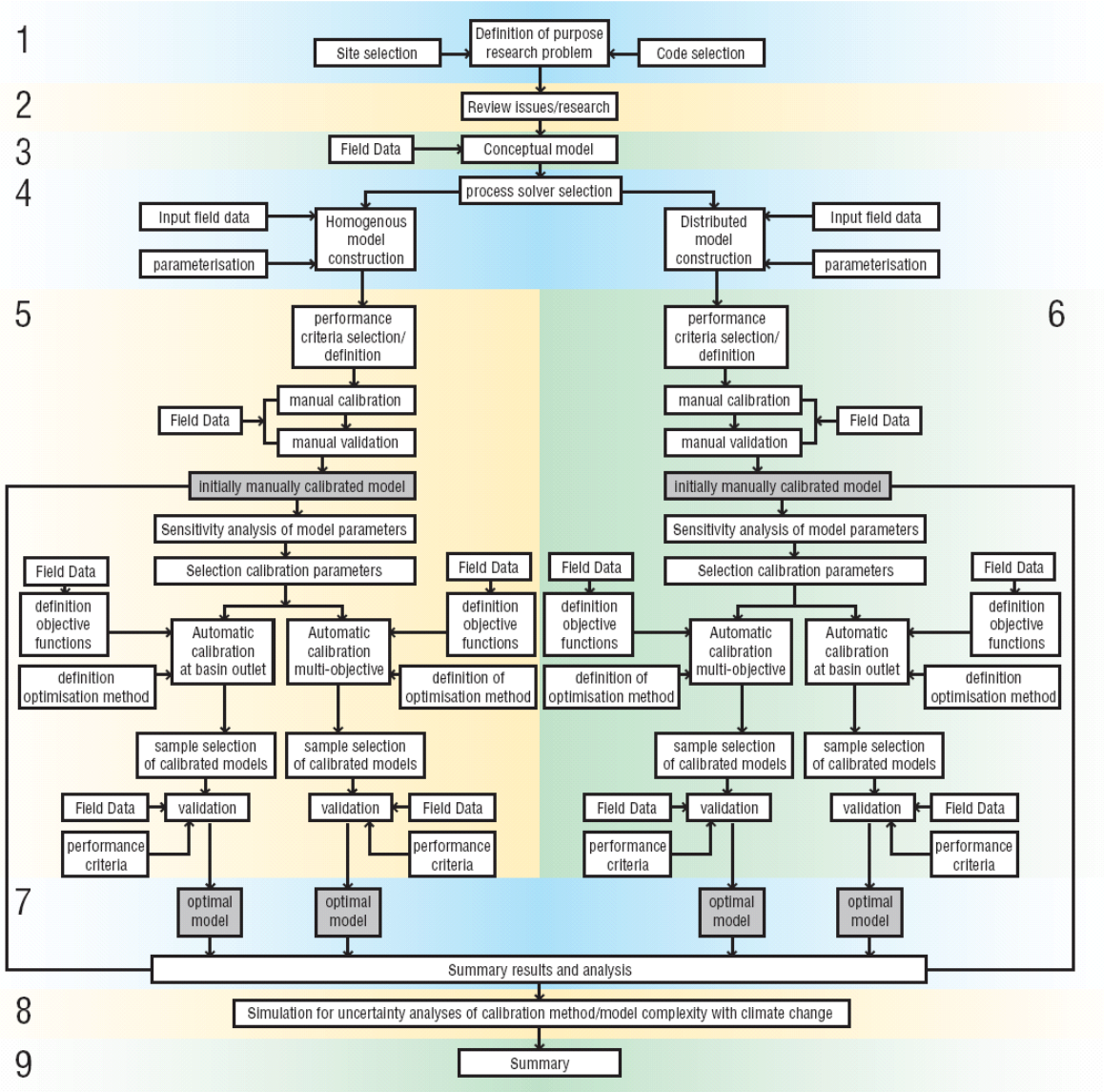


Figure 1.2. Conceptual research strategy to explore the thesis research questions and chapter division of thesis

1.4. Site and code selection

Figure 1.2 shows that the initial step is to define the purpose of the research which requires the selection and justification of the site and model code. In order to select a suitable catchment for the research, the following criteria were defined:

- A medium sized catchment of between 500 – 1000 km^2 that enables differences to be seen between the outputs of different spatial complexities of the proposed models. This size of catchment should also be typical of that in which water resources problems are often addressed at local to regional scales (e.g. Refsgaard, 1997)

- A catchment with substantial spatial variation in precipitation, topography, soils and/or geology, so that when different complexities of models are developed, scale differences may be expected to impact the results.

- A catchment that has been substantially surveyed and monitored with sufficient data to develop a distributed model. These data must also be freely, cheaply or routinely available for research purposes.

In order to undertake the large amount of modelling that is central to the thesis, the emphasis was on the synthesis of secondary data. Within the timescale of the research it was not feasible to undertake detailed primary data collection.

Given the criteria, initial research was undertaken and the Lowland Catchment Research (LOCAR) catchments in the UK were considered as feasible options. These catchments included the Frome/Piddle system in Dorset, the Pang/Lambourne catchments in Oxfordshire and the Tern catchment in Shropshire. The NERC funded LOCAR project (2000-2006) investigated how water enters, is stored within, and is discharged from rivers in these three groundwater-dominated catchments (NERC, 2008). Research into the three sites highlighted that the 876.36 km^2 Tern catchment in Shropshire, detailed further in Chapter 3, met all the criteria defined above. The catchment is an appropriate size and has a characteristically complex geology of permeable sandstones, less permeable mudstones and substantial glacial drift deposits. Its status as a LOCAR catchment means that extensive data (prerequisites for the development of a complex hydrological models) are available.

Different model codes were assessed, as documented further in Chapter 2, in order to identify one that would be able to meet the research requirements. It was necessary to select a model code that is able to operate with uniform parameter representation (a homogenous spatial distribution), as well as a more fully distributed spatial representation. In order to assess model performance at numerous gauging stations and groundwater level boreholes it is also necessary for the spatially homogenous model not to be a typical lumped model. It would be possible to construct different spatial complexities of model using different codes, the approach adopted by Refsgaard and Knudsen, (1996), but rather it is considered more appropriate to compare the effects of spatial complexity using the same modelling code so as not to introduce further inter-model code uncertainty.

As an objective of the research is to test different calibration methods, a further requirement of the modelling code was that it would be possible to undertake a range of calibration methods including both the facility to undertake manual calibration as well as automatic calibration methods.

The MIKE SHE modelling system (Abbott et al., 1986a &b) has been selected as a code which satisfies these criteria. As detailed in Section 2.4 and Chapter 4, the code can be used to create varying complexities of hydrological models from both conceptual as well as physically-based process representation. The MIKE ZERO framework in which MIKE SHE operates also has an additional component, Autocal, enabling automatic calibration of models.

1.5. Structure of the thesis

By exploring some of the key issues of uncertainty within the commonly used hydrological modelling framework, this thesis assesses the extent to which different choices in the construction and calibration process result in differences in simulated model outputs. Using a broad ensemble approach that employs a range of models developed using various modelling protocols, the research seeks to examine whether it is possible to develop a best protocol (defined by models of different spatial representations and calibration methods). Implications are assessed by quantifying the impacts and uncertainty of climate change, an expanding area of research to which hydrological models are applied.

The thesis is structured in a further eight chapters which broadly follow the stages of the conceptual research strategy shown in Figure 1.2. As shown in Figure 1.2, Chapter 2 reviews the background concepts and research concerned with key issues in hydrological modelling, data and uncertainty. The chapter also provides an overview of the MIKE SHE modelling code that is used in the research.

Chapter 3 provides a detailed site description and data analysis of the Tern catchment. It provides a conceptual understanding of the catchment and reviews the data that are subsequently used in the construction of hydrological models. Although using secondary data, the chapter provides a comprehensive review and includes a predominantly primary analysis of the data. The MIKE SHE hydrological model setups are detailed in Chapter 4. As shown in Figure 1.2. both a homogenous catchment model with predominantly uniform parameter values, and a spatially distributed model are developed. The two catchment models also include a coupled MIKE 11 hydraulic river model that is also described in this chapter.

Chapter 5 is the first of two chapters which presents modelling results. As shown in the conceptual strategy (Figure 1.2), the chapter includes details of the performance criteria, the initial testing and manual calibration of the homogenous model. A parameter sensitivity analysis is also included in order to define those parameters that are then taken forward and subject to automatic calibration. Two types of automatic calibration are reviewed, one that automatically calibrates river flow at the basin outlet, the other using a more rigorous multi-objective approach featuring river flow and groundwater levels at a number of locations. Figure 1.2 shows that Chapter 6 follows the same approach as Chapter 5, but with the focus on the manual calibration, sensitivity analysis and automatic calibration of the distributed model.

In Chapter 7, modelling results and uncertainty are reviewed with reference to the first five research questions outlined in Section 1.2. Chapter 8 seeks to put the main body of the research into context by addressing the last research questions concerned with model structure and calibration uncertainty when simulating the impacts of climate change.

Chapter 9 makes summaries of the principal conclusions of the research, highlighting the contribution the thesis makes to the research field. It also suggests future directions for further study.

Chapter 2

Hydrological modelling, data and uncertainty

2.1. Introduction

This chapter provides an overview of the background concepts and some of the main issues and uncertainties within hydrological modelling. Section 2.2 defines a classification of different types of hydrological model and compares some of the most widely used models. The issues and uncertainty associated to physically-based, distributed modelling are then discussed in Section 2.3, as MIKE SHE, the model code used in this research is classified as this type of model. An overview of the key processes and applications of MIKE SHE are then provided in Section 2.4. Section 2.5 briefly reviews additional error and uncertainty that derives from input data into hydrological models, and definitions, issues and methods of model calibration and validation are discussed in Section 2.6, where parameterisation, performance measures and sensitivity analyses are also discussed.

2.2. Classification and comparison of hydrological models

This section outlines the development of hydrological models, reviewing and comparing some of the most widely adopted hydrological models and modelling approaches. A classification of hydrological models is given according to three criteria; process, spatial and system representation.

2.2.1. System representation: deterministic and stochastic

There are two approaches used to model hydrological systems, deterministic and stochastic system representation (Shaw, 1994). Deterministic (or mechanistic) models seek to simulate the physical processes operating in the catchment, and are usually concerned with the conversion of volumes of water from precipitation to streamflow. The models identified in Table 2.1 are predominantly deterministic in nature, and MIKE SHE, the model code used in this thesis can be classified as a deterministic model.

Stochastic models such as the stochastic differential equation lumped rainfall-runoff model used by Lee et al., (2001) seeks to describe a hydrologic time series of measured variables such as evaporation and rainfall, and attempts to include the elements of probability that are intrinsic to earth systems (Shaw, 1994). In this case, Lee et al., (2001) treat the measured values of rainfall as the random element within the model to include a range of uncertainty that is associated in the measurement of rainfall data. The primary difference between deterministic and stochastic models is that deterministic models will always yield the same result, no matter how many times a simulation is run for a given setup. Stochastic models on the contrary will produce different results each time the model is run and are in essence 'non-repeatable' due to the probability of occurrence.

The majority of applied hydrological models in widespread use are deterministic in nature, although a more stochastic approach is encroaching into the research field with the use of deterministic models that are being driven with stochastically derived inputs such as can be derived within the Generalised Likelihood Uncertainty Estimation (GLUE) framework that is discussed further in Section 2.6.3.3 (Beven and Binley, 1992).

2.2.2. Process representation: empirical, conceptual & physically based models

The simplest type of hydrological models are 'empirical' or 'metric' which depend entirely on observations and field data. Empirical models attempt to represent relationships between input and output time series using transfer functions. Empirical models stem back to the unit hydrograph theory (Sherman, 1932) that considers the linear relationship between effective input precipitation, assumed to be uniform in space, and the runoff response in a given unit of time, e.g. one hour or one day. Simple

rainfall-runoff relationships which plot rainfall and streamflow can be simply derived for a catchment as in Figure 2.1. Law (1953) used data for the River Derwent catchment (Trent) from 1906-1947 to derive the relationships shown in Figure 2.1 where it is seen that the correlation coefficients between runoff and rainfall are 0.87 for the winter period and 0.91 for the summer period, which indicate a satisfactory linear relationship.

Despite their benefits, derived relationships are basin and usually time-period specific. For example, if land-use change has occurred within the catchment then the relationship may no longer be applicable, and its reliability questioned when extrapolating for an event outside of that from which the relationship was derived. However, these simple empirical models remain useful in establishing general catchment characteristics.

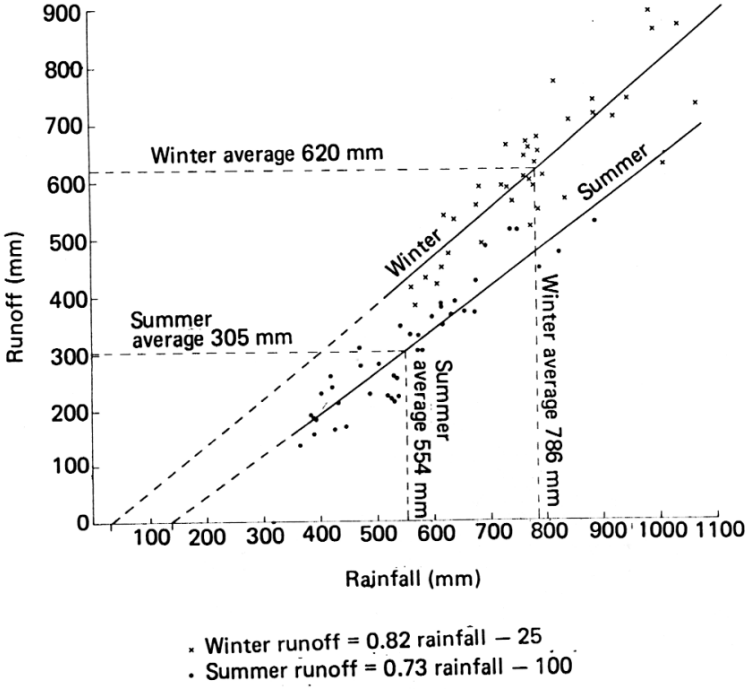


Figure 2.1 Example of rainfall runoff relationships that show average catchment characteristics (Law, 1953, cited in Shaw, 1994).

Later developments of empirical models include Artificial Neural Networks (ANNs) which use existing catchment data to learn the behaviour of the rainfall-runoff process by supervised or unsupervised pattern recognition between inputs (Caudill and Butler, 1992a; b).

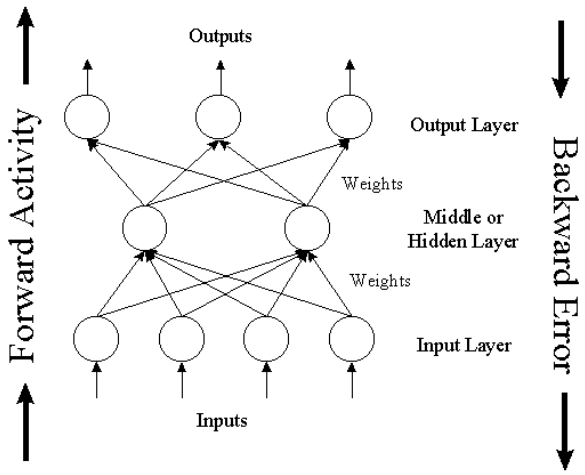


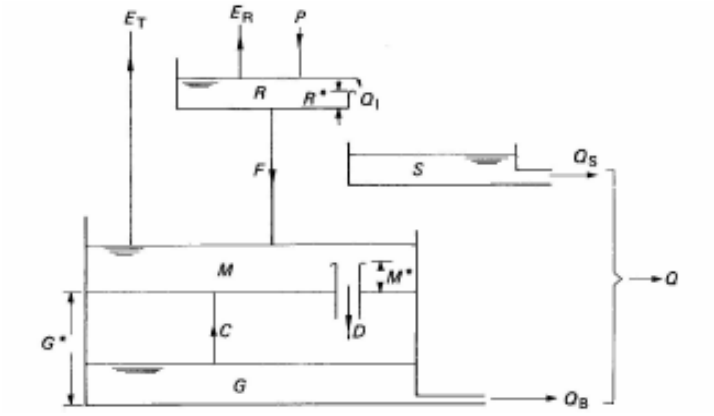
Figure 2.2. The typical Artificial Neural Network (ANN) conceptualisation (Caudill and Butler, 1992a)

ANNs use a three layer system as represented in Figure 2.2. A base layer requires inputs (e.g. precipitation). A second hidden neurode layer combines a number of the inputs and produces an output which is transmitted to many different locations (which can include other neurodes within the same layer). Connections between the input layer and the middle or hidden layer contain weights. The middle layer sums the weighted inputs to produce a single value output from each neurode. The sum is then used in a transfer function to create an output value in the third layer.

Maier and Dandy (1997) report that ANNs were first introduced to hydrological modelling by Daniell (1991) where an ANN was used to predict monthly water consumption and estimate flood occurrence. ANNs have been used for a variety of water resource applications including time-series prediction for rainfall forecasting (French et al., 1992) rainfall-runoff processes (Minns and Hall, 1996; Shamseldin, 1997) and representing soil and water processes including soil moisture (Altenford, 1992).

A second way to represent hydrological processes in a model is by using a conceptual approach. Conceptual models represent processes schematically – from precipitation input to streamflow output, with a series of stores with relatively simple mathematical equations describing the connecting fluxes. Water budgets are calculated for each of the stores (regarded as a series of reservoirs) where the volume of water held varies with each given time step (Shaw, 1994).

A classic example of a conceptual model is the O'Donnell model (Shaw, 1994) (Figure 2.3) which is constructed around four storages Rainfall (R), Channel Storage (S), Soil Moisture storage (M) and Groundwater storage (G). This generalised catchment model utilises inputs of precipitation and evaporation and is governed by nine control parameters which can be calibrated using a curve fitting method to achieve a best-fit to stream discharge.



Where Q (Stream discharge) is calculated by:

Parameter	Description
P	Precipitation
E_R	Evaporation
F	Infiltration
R	Store
R*	Threshold of R
Q_I	Surplus of R after which water can be made available for Q_s
Q_s	Surface runoff
Q_B	Baseflow
S	Surface store for water resulting from Q_I
E_T	Evapotranspiration that can occur from store M
M	Unsaturated zone storage for infiltrated water from F.
M*	Threshold above which D can occur
D	Percolation, only when water volume in M is greater than threshold M*
C	Capillary rise to store M, unless G attains the G* threshold
G	Groundwater store
G*	Threshold for G, above which G and M are combined

Figure 2.3. The O'Donnell Model as represented by Shaw (1994)

The equations describing many of the 'flows' within conceptual models do not have any true physical meaning (i.e. the parameter values cannot be acquired from field measurement). This results in the necessity to calibrate such models to observed data. It can be seen in Table 2.1 that the majority of hydrological models classified as conceptual are also often lumped or semi-distributed models (e.g. HYRROM, HBV, Stanford watershed model), as will be discussed in Section 2.2.4.

Thirdly, physically based models represent hydrological processes in a similar way to conceptual models in that they seek to isolate the component processes such as unsaturated and saturated zone flow, overland flow and evapotranspiration. However, instead of using parametric equations to represent these processes, physically based models use mathematical-physics using the equations of real world motion and quantum mechanics, hydro and thermodynamics (Beven, 2001). For example, the Institute of Hydrology Distributed Model (IHDM) (Beven et al., 1987) and MIKE SHE (Refsgaard and Storm, 1995) use physically based laws such as St Venant equations of channel flow and 3D finite difference groundwater flow governed by Darcian law to solve different components within the model. As Table 2.1 highlights the hydrological models that have been classified as physically-based, are also often classified as distributed models.

2.2.3. Spatial representation and grid descretisation: lumped, distributed and semi-distributed models

There is a further degree of classification within hydrological modelling that is concerned with spatial representation of the model. There are three primary spatial representations used: lumped, semi-distributed and distributed models (Figure 2.4). Lumped models (Figure 2.4.a) class the basin as a single unit with model parameters that attempt to represent the average values across the catchment (Refsgaard and Knudsen, 1996).

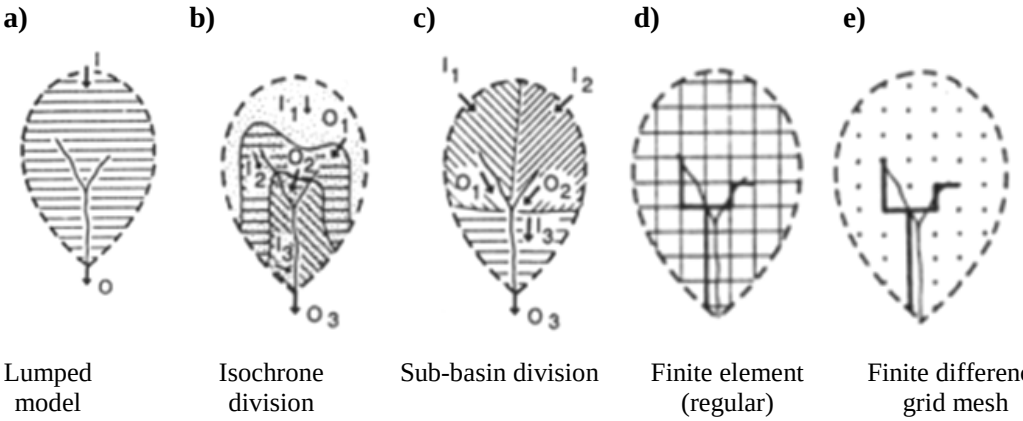


Figure 2.4. Conceptualised views of spatially representing the catchment with the lumped, semi-distributed and distributed approaches

Many of the earliest hydrological models such as the Stanford Watershed Model (Crawford and Linsley, 1966) are classified as lumped models. Most of the more recently developed models are often semi-distributed (Figure 2.4 b&c) or distributed in

nature (Figure 2.4 d&e), this could be attributed to the relationship between hydrological model development and the development of computer processing speed and storage capacity (Kirkby et al., 1993).

Distributed models (Figure 2.4.d&e) have the ability to take account of spatial variation of all parameters and variables within the catchment. Distributed models are usually physically based (Table 2.1) and descretise the basin into a network of grid cells or large number of elements, and solve the parameter equations for each of the grid cells independently (Beven, 2002). The finite difference/grid mesh method (Figure 2.4.e) is the method used within the MIKE SHE model, within a regular spatial grid, for representation of the surface and sub-surface flow equations (DHI-WE, 2005 and Beven, 2002). By contrast, the finite element method, shown in more detail in Figure 2.5 can be used to distribute data at varying resolutions across the catchment with areas around the catchment boundary represented with larger triangular elements than those closer to the river network. This method has also been widely used to represent elevation which is often represented in a Triangular Irregular Network (TIN).

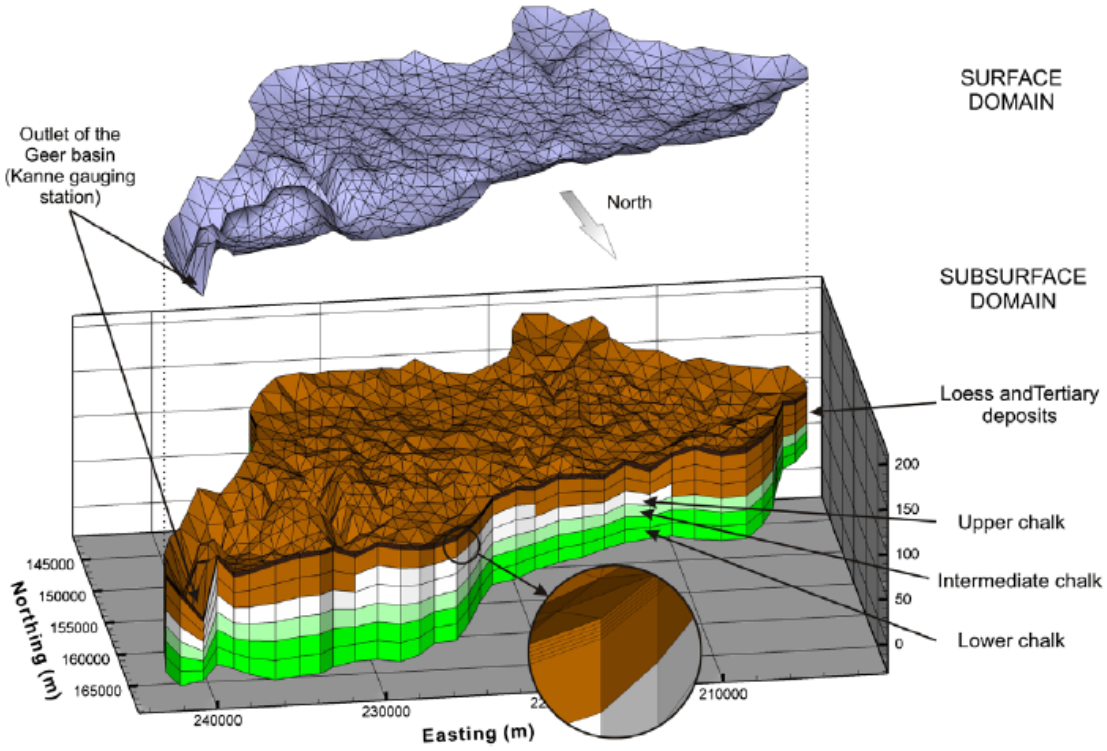


Figure 2.5. Conceptualisation of the spatial descretisation of the Geer basin using a finite element mesh (Goderniaux et al., 2009)

There are often hundreds to thousands of grid points or cells within a distributed model, with many parameters associated with each cell. This means that there can be up to two or three orders of magnitude more parameters in a distributed model compared to a lumped model covering the same area (Refsgaard, 1997). A schematic of the process of finite difference distributed modelling is given in Figure 2.6. This highlights how the output from each grid cell is available for input to the adjacent horizontal or vertical grid cells.

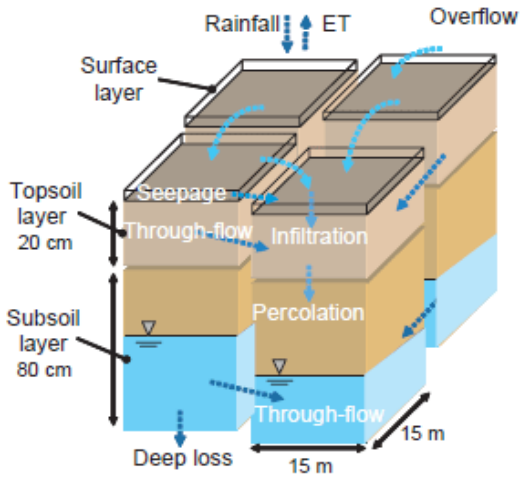


Figure 2.6. Example of the finite difference or grid method as a means of spatially distributing hydrological models (source: ss.jircas.affrc.go.jp/.../2004/2004_04.html)

Due to data availability issues, most distributed models are required to use average variables and parameters at the specified grid scale. Beven (2001) discusses the classification of distributed models with regard to the problem of acquiring suitable volumes of data – such models have consequently been regarded in a sense as lumped conceptual models at the element scale (Beven, 2001). Section 2.3 further discusses data and grid size issues within distributed modelling.

Further to lumped/uniform and distributed models, the principle of a semi-distributed model combines the idea of both a simple lumped model and a distributed model, as depicted in Figure 2.4 b&c. With a semi-distributed catchment model, the basin is represented by a number of separate zones, whether topographically divided (division of areas by elevation, e.g. upland or lowland) as with TOPMODEL (Beven and Kirkby, 1979), or by sub-basin distribution such as in Figure 2.4c.

Within a semi-distributed model, parameters representing the processes within each zone are independently calibrated to account for some of the spatial variation within the

catchment. Other approaches of specific models to sub-divide the watershed range from descretisation schemes based on the tributaries of the river network, for example in the DPHM-RS model used by Biftu and Gan (2001), to divisions of the catchment with Hydrologically Similar Units (HSUs) or Aggregated Simulation Areas (ASAs) in the SLURP model (Kite, 1995), or hydrological response units (HRUs) in the SWAT model (Arnold et al., 1993).

Kirkby et al., (1993) suggest that treating the hydrograph response in a number of partitions may be the best method. The suggested partitions of the unsaturated infiltration zone, saturated subsurface and surface zones, channel network and groundwater in aquifers are in general accordance with the Hydrologically Similar Unit (HSU) concept of Becker (1992; 1995), and Becker and Braun (1999). HSU aggregate areas of hydrologically similar behaviour into different profiles.

Figure 2.7 further illustrates the concept of sub-dividing the catchment into characteristic profiles (Karvonan et al., 1999). It is possible to have numerous profiles that can be replicated multiple times within the catchment. Each profile can be given its own parameterisation and calibration but all are linked to the overall input precipitation and losses from evapotranspiration and all are linked by channel routing to the catchment outflow. With the HSU approach, the separate profiles enable the modeller to consider the different processes that may be operating in the different profiles. For example, drainage flow may be important within an agricultural field profile and thus can be included. This may however be a process that is omitted from a forest profile in which subsurface flow and surface runoff are key processes.

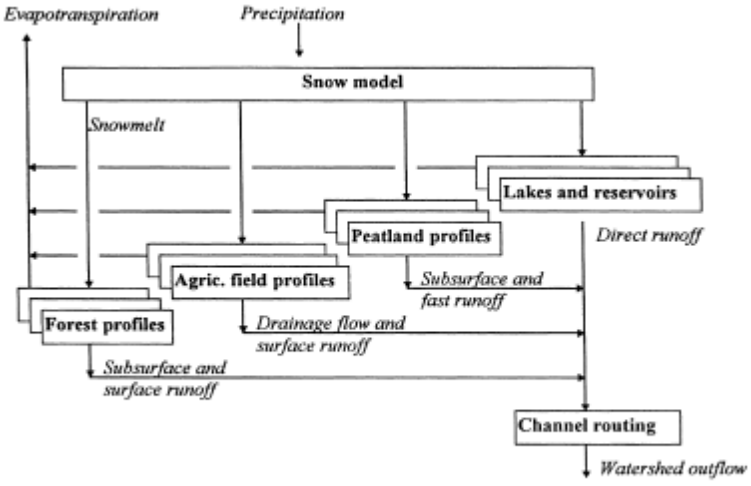


Figure 2.7. Hydrologically Similar Unit (HSU) concept as a method of semi-distributing a model (Karvonan et al., 1999)

In summary, physically-based distributed models are often considered as superior to simpler conceptual ones. In theory they are more exact and in principle they therefore require less calibrating and fitting of parameters. However, conceptual models are frequently the reasonable compromise when a model's demands for data meet the reality of application (Refsgaard, 1997). Complex solutions are not always warranted, and simple solutions can be just as adequate (Loague and Freeze, 1985). As pointed out by Refsgaard (1997), there may be no final conclusion on which type of model is better, but the most important factor is the proper consideration in choice of strategy and type of model that is appropriate for the question/issue to be solved.

2.2.4. Comparison of hydrological models

Having explained the theoretical classifications of hydrological models according to the three main criteria above, Table 2.1 seeks to place some of the most widely used models within this context. Although there are almost countless numbers of model codes (with a good review given by Singh et al., (1995)), the particular ones have been chosen to be representative of the different classifications outlined in the previous Section, 2.2.3. Table 2.1 shows that models have been constructed in many developed countries; predominantly by research institutes and universities. With the exception of empirical models, a key feature of hydrological models is that the same model structure or 'model code' (Andersen et al., 2001) can be used for a number of basins of varying size and location, with the parameters assigned values bespoke to each catchment.

It can be seen that some of the models such as the Stanford Watershed Model (1966) and HBV (1976) are quite old in comparison to others such as SLURP (1995) and MIKE SHE (1995) derived from SHE (1986). The older models often require less computing power and hard disk space for data storage, and are often classified as lumped and conceptual models as opposed to the more recent distributed and physically based models that are computer intensive.

Table 2.1 highlights example applications for the different models. It can be seen that physically-based and distributed models are frequently used in complex studies such as surface water – groundwater interaction research (Stadnyk et al., 2005), hydro-chemical routing in pollutant and water quality studies (Santhi et al., 2006) and river basin management (Jain et al., 1992).

Table 2.1. Review of a selection of widely used hydrological models

Name of Model	Attributed to/developed by	Spatial representation	Process representation	System representation	Data Requirements	Brief description	Applications
HYRROM Hydrological Rainfall-Runoff Model	Blackie and Eles, (1985) – UK Institute of Hydrology	Lumped	Conceptual	Deterministic	Precipitation	Flow simulated with simple representations of physical processes. Easy to use. Nine parameters available for calibration.	Infilling missing flow data. Quality control of data. Generating synthetic flow sequences. Water resources assessment.
SWM4 Stanford Watershed Model	Crawford and Linsley , (1966) - Stanford University	Lumped *(can be quasi-spatially variable)*	Quasi-physical but considered conceptual	Deterministic	Precipitation and potential evapotranspiration, radiation, temperature, cloud cover, wind, tide.	Uses a soil moisture accounting procedure and represents hydrological processes within the drainage basin through storage and routing functions. 34 parameters (25 without snowmelt) available for calibration.	Applied globally, predominantly for civil engineering design and agricultural engineering (Fleming 1975).
HBV	Bergström, (1976) and Bergström et al., (1992) – Swedish Meteorological & Hydrological Institute (SMHI)	Lumped *(can be modified to semi-distributed and elevation zones)*	Conceptual	Deterministic	Sub-basin division, altitude and land cover distribution, time series of precipitation and temperature (time-series of observed water discharge at some sites).	Originally a forecasting and simulation tool. Daily rainfall-runoff model with conceptual numerical descriptions of hydrological processes at catchment scale.	Used in over 40 countries. Flood forecasting in Nordic countries. Also for spillway design flood simulation (Bergström et al., 1992), water resources evaluation (Jutman, 1992) & effect of wetlands on regional nitrogen transport (Arheimer et al., 1998).
TOPMODEL	Beven and Kirkby (1979) – University of Leeds, UK.	Semi-distributed - subdivision into small homogenous sub-basin units modelled	Physically-based	Deterministic, but can be run stochastically	Topographic data, limited soil data, other parameters from direct measurement.	Collection of concepts that can be used and adapted to specific study. Combines spatial variability of source-areas with average response of basin soil-	Global & multidisciplinary. Early climate change scenario testing (Wolock and Hornberger, 1991). Recently modified

TOPMODEL *(cont).*		separately.				water storage (reducing the number of parameters). Flow estimated with assumption water table follows topography (a surrogate for hydraulic gradient).	version for stream chloride concentrations (Page et al., 2007). Vast literature of non-commercial applications (uncertainty estimation), (Xavier et al., 2006) and (Freer et al., 1996).
SLURP Semi-distributed Land Use-based Runoff Processes	Kite, (1995) - National Hydrology Research Institute, Canada	Semi-distributed - divides the watershed into hydrologically-consistent sub-units known as aggregated simulation areas	Conceptual (quasi physical)	Deterministic	Topographic data, land cover data, climate and hydrometric data.	Continuous simulation model. Parameters related to land cover (vegetation type). Most important parameters in model include interception coefficients, depression storage, surface roughness, infiltration coefficient, groundwater conductivity and snowmelt rates. Model accounts for changes in the distribution &type of land cover over time. Suitable for climate change impact studies.	Applications vary in size of catchment. e.g. Prairie sloughs, of few hectares (Su et al., 1997) to 1.8 million km^2 Mackenzie basin (Kite et al., 1994). Range of applications from hydro power production (Kite, 1997), estimation of irrigation performance (Kite and Droogers 1999). Woo and Thorne (2006) assessed snowmelt contribution to runoff in mountain catchment, Canada.
SWAT Soil and Water Assessment Tool	(Arnold et al., 1993, 1998) United States Department of Agriculture – Agriculture Research Service	Semi-distributed (HRUs) – grid descretisation at users choice.	Physically-based	Deterministic	Multiple inputs ranging from precipitation, temperature, solar radiation, wind speed, PET, land cover, elevation, fertiliser. Available input at varying descretisations (Neitsch et al., 2005).	River basin scale model developed to quantify impact of land management practices in large, complex watersheds. Daily time step. Divides catchment into HRUs where sub-basins have homogenous climate, soil, management and land cover.	Kang et al., (2006) – Applying SWAT total nitrogen, phosphorus and suspended solids in small rice paddy catchment, Korea.

SWAT *(cont).*						Can predict effects of watershed management on runoff, sediment, nutrients and pesticide yields.	Santhi et al., (2006) studied impacts of water quality management plans implemented at watershed and farm level in Texas.
WATFLOOD	(Kouwen et al., 1993) University of Waterloo, Canada	Distributed (according to input raster satellite data).	Physically-based	Deterministic	Radar rainfall data, LANDSAT or SPOT land use and/or land cover data. Gauged precipitation for infilling and calibration, flow, snow depth, temperature and radiation.	Flood forecasting and long-term simulation using distributed precipitation data from radar or numerical weather models. Satellite data directly incorporated into model. Multiple processes modelled including interception, infiltration, snow accumulation and ablation, recharge, base flow.	Pietroniro et al., (2006), Toth et al., (2006) for climate change impact assessments. Groundwater separation study in boreal wetland terrain Stadnyk et al., (2005).
MIKE SHE Derived from System Hydrologique Europeen	(Refsgaard and Storm, 1995) – DHI Water and Environment (DHI-WE)	Lumped, semi-distributed or Distributed	Physically-based (with some components optional conceptual approach)	Deterministic but with ability to run stochastically using a Monte-Carlo autocalibration method.	Various parameter data for: Topography, Precipitation, temperature, land use, evapotranspiration, overland flow, unsaturated zone flow, saturated zone flow, groundwater.	Modular structure comprising six process-orientated components representing physical processes of land phase of hydrological cycle. Data is input discretely in a horizontal orthogonal network of grid squares so that the parameters can be represented at a high spatial resolution.	See Table 2.11 for more detailed review of MIKE SHE applications. Applications include: Modelling groundwater Christiansen et al., (2004), Pollution and Water quality (Brun, 2002a), Irrigation Jayatilaka et al., (1998), River Basin Management Jain et al., (1992).

By aggregating the models and placing them in direct comparison it may also be appreciated that often models with similar classifications (e.g. SLURP and SWAT, and WATFLOOD and MIKE SHE) have similar applications and similar data requirements. Studies have shown that regardless of the specific model code used, very similar results for the same question can be achieved. Some of the initial comparisons between hydrological model capabilities were carried out by Loague and Freeze (1985) using two simple black box and a quasi-physically based model. The models were applied to three small (<10km^2 catchments) where it was shown that 'despite poor performance' there were no significant differences between the different model results.

The applications given in Table 2.1 are only a very small representation of how each of the models have been applied. It can be noted that many of the given examples are for use in water resources planning or impact assessments of given changes such as climate change or land use change.

In addition to applied problem solving and management studies where hydrological models are used for scenario testing or impact assessment, a widening area of research that addresses model theory, uncertainty estimation and model development is also prevailing. For example, it is highlighted with TOPMODEL which has been used significantly for uncertainty estimations of input data, model grid resolutions (Chaubey et al., 2005) and sensitivity analyses to model parameterisations. To exemplify this issue further, Section 2.4.3.1 includes a review of 'research studies' within MIKE SHE.

2.3. Physically-based, distributed hydrological modelling

This section includes a discussion of four main issues associated with physically-based distributed hydrological modelling. A review is given of the issues concerning the data intensive nature of distributed models and requirement of abundant data sets. This is followed by the problems of representing parameters by homogenising them at a grid cell scale. The uncertainty in using physically-based equations at the grid cell scale is also described, where equations often derived at the micro-scale are up-scaled and used to represent often large grid cells. The issue of input data resolution and problem of choosing a suitable grid size in the modelling process is also discussed.

2.3.1. The requirement of abundant datasets

Great concern is frequently raised about the large amounts of data that distributed models such as MIKE SHE require. Many measurements of critical model parameters do not exist, or are not available at a high enough spatial resolution to render a model fully distributed. For example, input precipitation data distributed by Thiessen polygons from point/gauged data are able to represent some of the spatial variability of precipitation, but due to the often fragmented rain gauge network (discussed in more detail in Section 2.5.1) some of the spatial variability is inherently lost. Consequently there is a lag between the capabilities in distributed modelling and that of obtaining field data (Dunne, 1983; Silberstein 2006) as spatially distributed models are often capable of data input at finer spatial scales than typically used.

With the widespread development of remote sensing and GIS however, gaps in data are being frequently overcome. For example, Andersen et al., (2002a) and more recently Stisen et al., (2008) used METEOSAT based Leaf Area Index (LAI) measurements and AVHRR precipitation data in sub-catchments of the Senegal river basin in the MIKE SHE model to show how in an otherwise data sparse basin, marked improvements in model performance could be made with the addition of the spatial resolution of data made available by remote sensing.

2.3.2. Homogenising parameters at the grid cell scale

Another concern in distributed modelling is that the heterogeneity of processes occurring at the sub-grid scale are not accounted for when up-scaling to grid cells, especially those frequently set at low resolutions and for large catchments such as the 4km x 4km grid in the data sparse Senegal catchment (Andersen et al., 2002a).

As outlined above, a lack of hydrological data results in the need to compromise with lower spatial resolution and fewer grid squares, meaning the area of each grid square increases. As a result of this scaling, any given grid cell within a model is assumed to be homogenous as only one set of parameter values can be assigned to each grid cell. However in reality this homogeneity is not present. As detailed by Beven (1989) for example, overland flow must be lumped over the specified grid cell, when in reality this would vary internally within the area covered by the cell. In addition, other hydrological

processes would naturally vary within a grid cell such as precipitation variation that often varies considerably over short distances (Hutchinson, 1970). It is concluded that there is no given theoretical framework for lumping these sub-grid processes for spatially heterogeneous grid-squares and by doing so there is a 'conceptual-leap' introduced to distributed modelling (Beven, 1989).

2.3.3. Use of physically-based equations at the grid scale

A further issue in distributed modelling is that the physically-based equations that are used to represent each grid cell within a model are debatably not accurate enough to be used at the adopted catchment scale study, since the process equations that distributed models use have often been derived from micro scale studies (Beven, 1996). For example, Cosby et al., (1984) discuss the hydraulic properties of soil with regard to their relationship with soil texture. The research that was carried out was essentially physically-based using up to 5000 soil samples of 'fist-sized fragments' in a laboratory. Although comprehensive at the scale in which the study was undertaken, it does not consider the effects of larger scale processes also important in hydraulic conductivity such as macropore flow that are likely to operate at the grid scales used in hydrological modelling.

In spite of the criticisms, there are numerous studies that have shown an improvement in model results from the use of distributed models which even when accounting for the issues above, are 'more physically-based' than the older generation of lumped conceptual models (Beven, 1989). Michaud and Soorishan (1994) compared three hydrological models in a 150km^2 semi-arid catchment. The study compared the physically-based distributed model KINEROS of Woolhiser et al., (1990), the SCS conceptual lumped model and the SCS conceptual distributed model with eight sub-basins (McCuen, 1982).

Table 2.2 indicates the lumped conceptual model gave the least adequate results. Results from the distributed models indicated that when calibrated, both physically-based and conceptual gave similar results with RMSE of time to peak 36.1% and 30.6% respectively (of mean observed). Without calibration, however, the physically-based model simulated flow more sufficiently as shown in all the test statistics in Table 2.2 for peak flow, time to peak and volume.

Table 2.2. Validation statistics for Peak flow, Volume and Time to Peak (Michaud and Sorooshian, 1994)

	Peak Flow		Time to Peak		Volume	
	RMSE (Percent of Mean Observed)	Average Bias (Percent of Mean Observed)	RMSE (Percent of Mean Observed)	Average Bias (Percent of Mean Observed)	RMSE (Percent of Mean Observed)	Average Bias (Percent of Mean Observed)
			Calibrated			
Lumped SCS	99.2%	−52.7%	36.1%	+14.7%	104.4%	−16.9%
Distributed SCS	85.4%	+10.3%	19.9%	+10.3%	83.5%	−27.6%
KINEROS	89.9%	+3.7%	30.6%	−6.9%	73.1%	−26.6%
			Uncalibrated			
Lumped SCS	120.3%	−97.2%	111.7%	−85.5%	115.6%	−93.2%
Distributed SCS	113.0%	−68.0%	46.7%	−1.6%	109.5%	−68.9%
KINEROS	79.0%	−26.4%	28.6%	+21.6%	70.1%	−4.5%

2.3.4. The issue of input data resolution/ grid size

Another key issue that has to be considered in distributed hydrological modelling is how the behaviour of simulated runoff varies with increases in the size of the catchment and the dimensions of the individual grid cells within the model. Early theory of scaling in hydrological modelling suggested that at small and micro scales, spatial variability in topography, soils and precipitation govern the production of runoff (Wood et al., 1988). As scale and grid cells increase within the model, a wider range of variability of model parameters are contained within each sampled area. The premise follows that at a given 'large scale' all cells will yield almost identical responses and can be considered homogenous.

There is a choice to be made therefore, regarding the required and optimal resolution of input data. Spatial variability can be better accounted for with higher resolution data, but it is frequently not possible or impractical to include such high resolution data particularly if a catchment is primarily un-gauged or very large. Despite a better representation of variability within a catchment, data of very high spatial resolution can contain redundant information due to replication of values over very small distances and leads to an increase in required storage capacity and computer time (Omer et al., 2003). The trade off is thus to ascertain an adequate grid size where the homogeneity balances with the preservation of the characteristics of a higher resolution input field (Molnar and Julien, 2000; Sivapalan and Kalma, 1995; Bloschl and Sivapalan, 1995).

Figure 2.8 is used to demonstrate the issue of representing the catchment with different grid sizes using the MIKE SHE model code. It is shown that the topography representation on the 500m × 500m grid is coarse. This may influence the timing of the routing of water through in the model. If the cells are larger then adjacent movement between cells will take place more rapidly through the time steps than the smaller grid size where water has to move between more cells. The boundary of the catchment is also influenced with the different grid sizes where a better definition of the catchment shape is shown with higher resolution (smaller) grid sizes. The complexity and length of tributaries are also influenced by the different grid sizes, with larger grids resulting in more jagged representation of water movement.

As summarised in Table 2.3, considerable research has been undertaken regarding gridding in hydrological modelling using numerous model codes on catchments of various sizes and different climatic regions. These studies seek to ascertain the optimal grid size or best way to form, at a given scale, an average hydrologic response which has minimal variation in response as the catchment area increases (Wood et al., 1988).

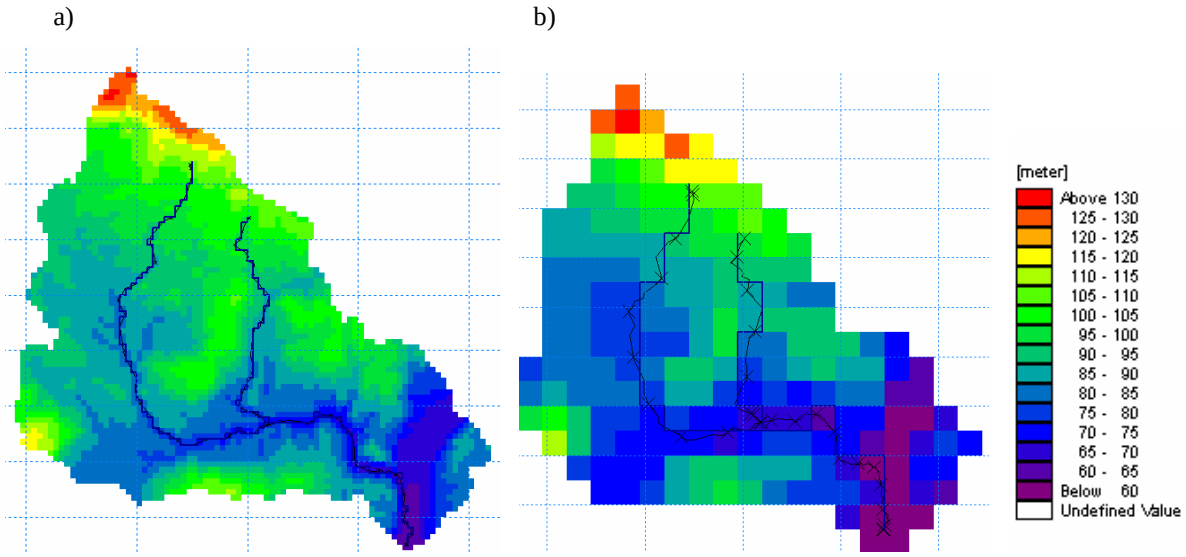


Figure 2.8. Bailey Brook, Shropshire, represented with a) 100m × 100m grid and b) 500m × 500m grid

Table 2.3. Review of studies associated with distributed hydrological model grid cell resolution

Author	Catchment location	Catchment size (km^2)	Grid sizes or number of sub-basins studied	Model used	Topic of study and main findings
Bathurst, (1986)	Wye catchment, UK.	10.5	250m, 500m	SHE	To divide catchment into elements no larger than 1% of total area
Wood et al.,, (1988)	Coweeta, N. Carolina, US	17.0	87, 39, 19, 3 semi-distributed sub catchments	TOPMODEL	Representative Elementary Area (REA): the smallest discernable point which is representative of the continuum. Strongly influenced by topography. REA = ~1km^2 in this study.
Zhang&Montgomery (1994)	Mettman Ridge, Oregon. Tennessee Valley, California.	0.3 1.2	2, 4, 10, 30, 90m	TOPMODEL	The need to compromise between increase in spatial resolution and data handling requirement. The highest grid resolutions finer than 10m do not give improved results.
Horrit and Bates, (2001)	River Severn, UK.	60 km length	10, 20, 50, 100, 250, 500, 1000m	LISFLOOD-FP	Studies inundated flood area and flood wave travel times. This study uses other hydrological variables than just runoff to give same conclusion that there is no improvement after100m when finer resolutions used. Shows that grid size affects shoreline location, flow routing behaviour & floodplain storage capacity.
Saulnier et al., (1997)	Maurets, France	8.4	40, 60, 80, 100, 120m	TOPMODEL	Grid size is analytically linked to the saturated hydraulic conductivity parameter (following Franchini et al., 1996). It shows there are scale dependencies on calibrated parameters and that model results and parameter values are not independent of scale.
Chaubey et al., (2005)	Moores Creek, Arkansas, US.	18.9	30, 100, 150, 200, 300, 500, 1000m	SWAT	DEM resolution affects delineation of the watershed and stream network. A decrease in spatial resolution results in a decrease in the volume of simulated stream flow. 100-200m required to achieve less than 10% error in flow simulation.
Molnar and Julien, (2000)	Goodwin Creek, Mississippi, US. Hickahala-Senatobia, US.	21.0 560.0	127 to 914m	CASC2D	Coarser grid resolutions can be used if the parameters are appropriately calibrated for each simulation. Found for both smaller and larger catchment.
Shrestha et al., (2006)	Huaihe, China (& sub-catchments) Wangjiaba Suiping	132, 350.0 29, 844.0 2093.0	10 minutes to 2.5◦	MASCOD	Development of IC Ratio (ratio between input resolution and catchment area). Similar results found for all 3 sizes of catchment. Model performance is more pronounced below 1:10, and rate of improvement negligible above 1:20. Optimal range is between this.

2.4. Overview of MIKE SHE

This section discusses in detail the MIKE SHE hydrological model code that is used to develop the hydrological models later in the thesis. An outline of the MIKE SHE code development from the SHE model is given and is followed by a discussion of how each of the components of the hydrological cycle can be represented within the code and summarises the equations that can be chosen for each of the main processes. A review of some of the applications of the model are also provided, demonstrating how MIKE SHE has both been applied at varying scales and complexities at many sites and scales around the world and in different climatic regions.

2.4.1. MIKE SHE origins and development

The Systeme Hydrologique Europeen (SHE) was a physically-based, distributed catchment modelling system (Abbott et al., 1986a, b). The development of the system was funded by the European Commission and developed by three European organizations from the UK (Institute of Hydrology), Denmark (Danish Hydrologic Institute (DHI) now DHI-Water and Environment) and SOGREAH, a French environmental consultancy firm. The prototype emerged in 1981, with the first production version in 1982. Distributed models such as SHE were developed in response to the identified inadequacy and inappropriateness of many conceptual, lumped rainfall-runoff models to simulate sub-catchment hydrological issues such as land-use change. The development of SHE, which uses real-world physical equations and allows for spatial variations within the catchment (e.g. topography, precipitation, soil and geology), resulted in the construction of a modelling concept where these issues could be addressed. The SHE model was then further developed in the 1980s by DHI-WE, formerly DHI in Denmark where the model was established as MIKE SHE, with the ability to represent more grid squares where as the original SHE was limited.

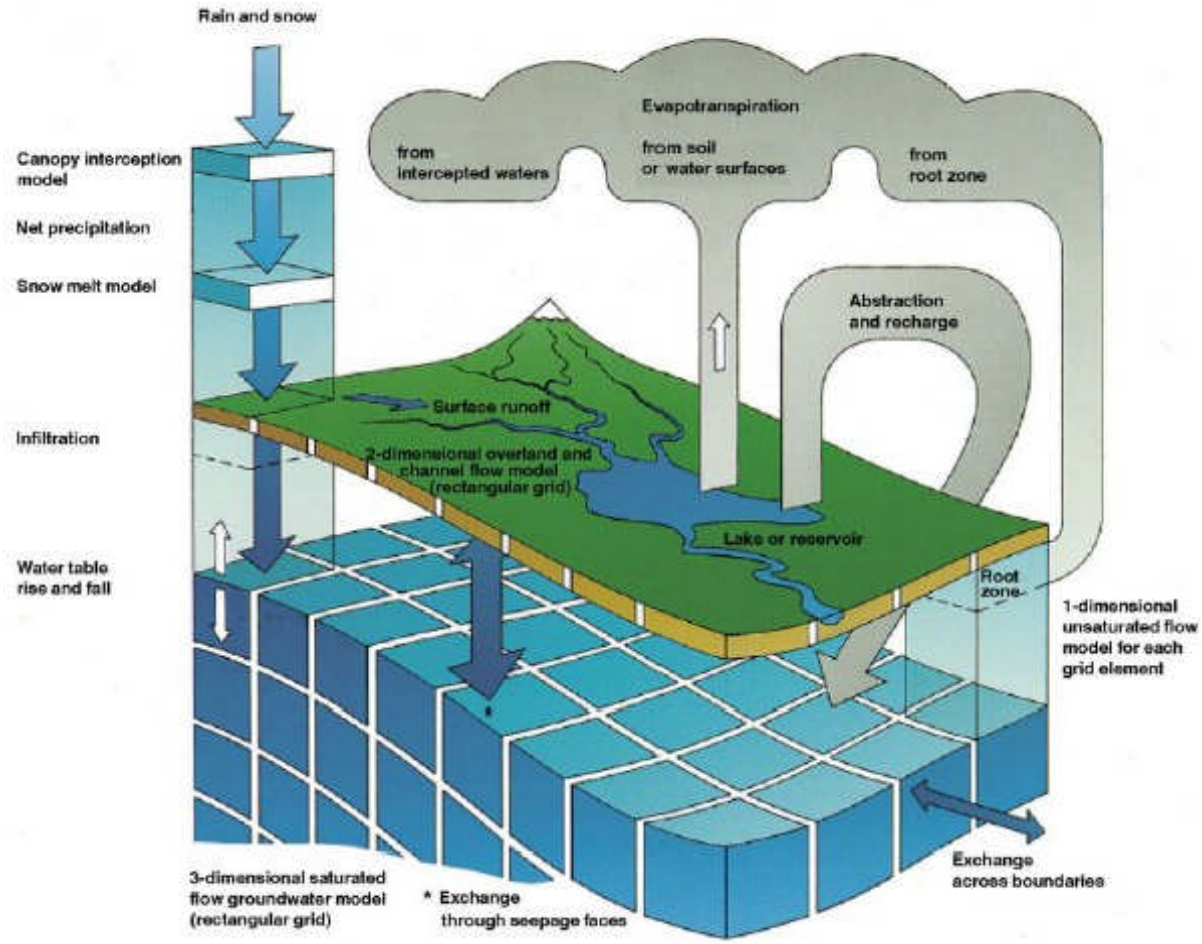


Figure 2.9. Conceptualisation of the distributed structure of MIKE SHE (Refsgaard and Storm, 1995).

MIKE SHE is a distributed physically-based code with an integrated description of the six process-orientated components that describe the major physical processes of the land phase of the hydrological cycle (Refsgaard and Storm, 1995). As conceptualised in Figure 2.9, the spatial distribution of catchment parameters are achieved in the horizontal domain by representation of the catchment by an orthogonal grid network and in the vertical by a column of horizontal layers at each grid square (Abbott, 1986a).

MIKE SHE includes a fully developed user interface where it is possible to select different solution techniques at various complexities for each of the hydrological components. The model code includes pre-processing and post-processing modules including analysis statistic options for comparisons in the calibration phase, and various options for displaying results. Various file formats can be used in model input, ranging from ESRI shapefiles to more simple text files which can then be edited and created into specific MIKE SHE files.

Table 2.4. Recent developments of MIKE SHE in the 2008 edition (summarised from dhigroup.com)

Component	Development	Details
MIKE SHE	Water Quality	The ability to simulate water quality across all the hydrologic cycle - including advanced sorption and degradation of dissolved solutes;
MIKE SHE	Elevation corrected precipitation with additional snow store/melt model	An improved, distributed rainfall-runoff model, with elevation corrected precipitation and rigorous snow storage and snowmelt module;
MIKE SHE	Unsaturated Zone / Macropore flow	Improvements to the Unsaturated Zone (UZ) module, including a physically-based macro pore model;
MIKE SHE	GUI	A series of user interface improvements, including the consolidation of climate input data and better access and visualization of pre-processed data.
MIKE 11	MIKE 11 GIS	The possibility to use ArcMAP to calculate the difference in results between two runs. A linkage between point source layers in ArcMAP and point pollutant loads in MIKE 11

The full detail and manual of the MIKE SHE code is given in the 'user's guide' (DHI-WE, 2005) which details all the system options and technical users reference with the detail concerning the governing equations. Refsgaard and Storm (1995) have provided a frequently referenced and detailed description of the structure and set up of the MIKE SHE model, including developments and changes made to the model code up to the date of publication.

A more recent synopsis of the model in its present form is available in Graham and Butts (2005). As MIKE SHE is a dynamic system which is continually evolving, further recent developments have been included in the 2008 edition of the model, these are summarised in Table 2.4 from information from the DHI website (dhigroup.com).

2.4.2. Process representation

As already introduced, it is possible to use different solution techniques to represent the different processes in MIKE SHE. For example, if the aim of the modelling work is to simulate groundwater (Madsen and Kristensen, 2002) then a more complex groundwater solution can be applied. Similarly, if the objective is to construct a large scale model (Andersen et al., 2001) then using the more complex solvers is often too demanding on time and resources for it to be viable or useful. For example, McMichael et al., (2006) used a simplified version of the MIKE SHE model where the model is set to use more of the conceptual equations rather than the physically-based ones. This

section provides an overview of how each of the different processes can be represented as well as the choice in numeric engines that can be used. The user options are expressed in Figure 2.10 as reviewed by Graham and Butts (2005). The figure aids in demonstrating the number of options the user has when setting up a MIKE SHE model.

2.4.2.1. Precipitation

In MIKE SHE precipitation is an input parameter into a model. Real data as a constant or a time series can be specified for either uniform distribution, station based using Thiessen polygons for example, or with full gridded spatial distribution. Precipitation is then represented for each grid within the distributed model based of the model domain and grid size which is preset. Additionally, air temperature can be input in the same format if snow cover and snowmelt are of importance in the catchment. Snowmelt is specified by a simple relationship where a given value (mm) will melt per day after temperature has risen above a specified threshold. As Table 2.4 indicates, the snowmelt component of MIKE SHE has been further developed for the 2008 release to include a more rigorous snowmelt and snow store model.

Interception of rainfall in MIKE SHE depends on vegetation type and stage of development (Poole et al., 1981), which is characterised by the leaf area index (LAI), where greater than 70% of rainfall is returned to the atmosphere via evapotranspiration.

2.4.2.2. Evapotranspiration

Evapotranspiration refers to the sum of the processes of direct evaporation from free water surfaces and transpiration from sub-surface water either directly or via plants (Graham and Butts, 2005). MIKE SHE calculates actual evapotranspiration (AET) using one of four options as indicated in Figure 2.10.

- *Soil Vegetation Atmosphere Transfer (SVAT)* – a model that comprises a single, semi-transparent canopy layer located above the soil layer. The only way for water to enter or leave soil is through this canopy layer.

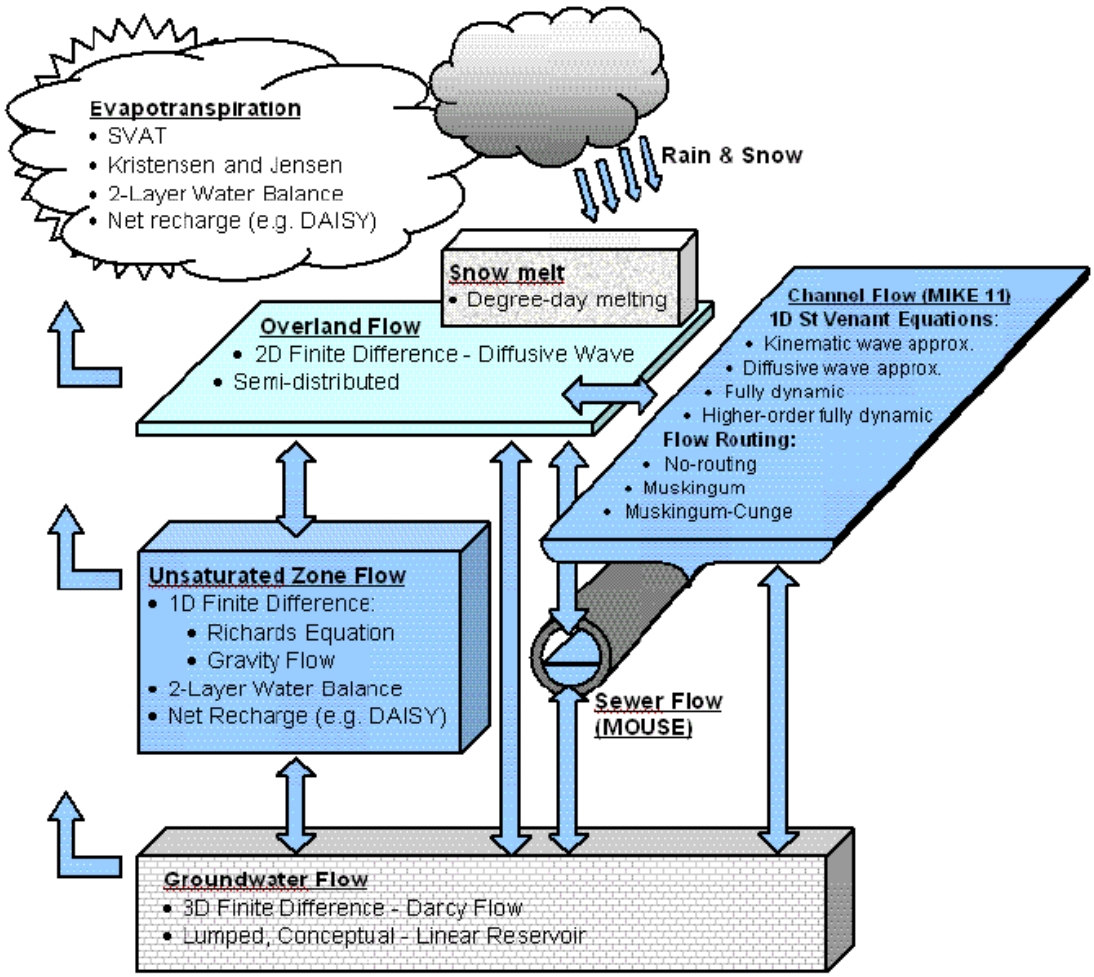


Figure 2.10. Schematic view of the processes in MIKE SHE. Available numeric engines are shown for each processes as well as the available exchange pathways for water between the process models. Source. Graham and Butts (2005).

- *Kristensen and Jensen Method* – using empirically derived equations Kristensen and Jensen, (1975) derived from fieldwork in Denmark. Time series for reference potential evapotranspiration (PET), leaf area index (LAI) and root depth (RD) as well as other empirical parameters that control the distribution of ET in the model domain and required as inputs for this method (Refsgaard and Storm, 1995)

- *Two-layer water balance method* – a simplified water balance method for the UZ storage and ET. Root depth defines the part of the UZ in which ET can occur. The method requires the same input data as Kristensen and Jensen method but is different in that it does not consider flow dynamics and so is better suited to areas where the water table is close to the surface such as in wetlands.

- *Open software systems - (e.g.* Daisy model). The Daisy model (Hansen et al., 1990) can be used as an added component to MIKE SHE to calculate evapotranspiration for studies especially concerned with agricultural studies such as Boegh et al., (2004). The model is a 1-D agro-ecosystem that can simulate crop production, crop yield, water and nitrogen dynamics (Abrahamsen and Hansen, 2000). Potential evapotranspiration data is used primarily, along with precipitation and irrigation inputs and other weather data to ascertain the upper limits for simulated evapotranspiration.

2.4.2.3. Overland (OL) flow

Overland flow such as that which is derived from ponded water as a result of precipitation that does not infiltrate into the UZ can be calculated by the two methods illustrated in Figure 2.10; the 2D finite difference method which is a diffusive wave approximation of the Saint Venant equations, as well as the semi-distributed, slope-zone approach. The finite difference method requires three parameters to be specified in the model: the Manning number, detention storage and the initial water depth (as detailed in Table 2.5).

Table 2.5. Parameters and conditions used in OLF calculation in MIKE SHE.

Parameter	Conditions	Description
Manning number	Overland Flow & Finite Difference	Manning M (equivalent to Stickler roughness coefficient). Manning M is the inverse of Manning n. Typical values between 100 and 10 with lower values used for OLF compared to channel flow.
Detention storage	Overland Flow & Finite Difference	To limit the amount of water that can flow over the ground surface (i.e. surface water must exceed this threshold before available to OLF). Water held in detention storage is available for infiltration to UZ and to evapotranspiration.
Initial water depth	Overland Flow & Finite Difference	Initial water depth at start of simulation held on the ground surface (ponded water).
Overland-Groundwater leakage coefficient	Overland Flow & Finite Difference AND reduced contact in subareas	For when the soil profile is saturated and the UZ is disabled. If detention storage is operating then there is direct flow between OLF and SZ. Vertical hydraulic conductivity in SZ if often not representative of permeability in the top layer of the soil and hence a leakage coefficient can be specified. Often used under permanently flooded lakes and floodplains where fine sediment may have accumulated forming a top layer.
Separated flow areas	Overland Flow & Finite Difference AND separated flow areas	To divide the model into OLF zones (conceptually these are areas separated by dykes/embankments). Flow between zones is prohibited.
Overland flow zones	Overland Flow & sub catchment-based OLF	Used to define topographic zones for the simple, catchment based OLF solution. Zones are defined as areas with similar topography. Additional parameters must be specified: the slope, slope length, Manning number, detention storage and initial depth.

The semi-distributed approach is similar to that of the Stanford Watershed Model (SWM) (Table 2.1), and is based on an empirical, conceptual relationship between flow depth and surface detention and uses the Manning equation of surface roughness.

2.4.2.4. Unsaturated Zone (UZ) flow

MIKE SHE simulates unsaturated zone flow vertically assuming gravity is the most important process in governing directional movement of water. Although lateral movement may occur especially in more hilly areas, it is ignored as it is computationally expensive to calculate. UZ flow is characterised by fluctuations in precipitation and hence soil moisture, as well as evapotranspiration and recharge to groundwater.

If the full Richards equation is used, UZ flow calculations are often very time consuming. As a result it is possible to lump UZ flow calculations by solving all the equations once and then applying the results to similar cells, for example, where the same precipitation and soil types are located. Table 2.6 reviews the methods available within MIKE SHE of calculating flow in the UZ.

Table 2.6. Description of the different engines available in MIKE SHE to compute UZ flow

Method	Description	Requirement	When to use
Richards equation	Physically-based on continuity equation and Darcy's law. Vertical movement driven by gradient of hydraulic head	- Pressure head as a function of saturation (moisture retention curve) - Hydraulic conductivity	Most computationally intensive but most accurate. Use for study of UZ flow dynamics
Gravity flow procedure	Assumes a uniform, vertical unit-gradient and ignores capillarity. A simplification of the Richards equation. Pressure head is ignored and vertical flow assumed from gravity. Faster and more computationally stable than the full Richards equation.	- Hydraulic conductivity for saturation relationship	When interested in time varying recharge to GW based on actual ppt and AET, not the dynamics in the UZ. (such as McMichael et al., 2006). Also for coarse soils when capillary pressure is small
Two-layer water balance	Conceptually divides UZ into a root zone and zone below the root zone. The method calculates the volume of water that will recharge the saturated zone.	- Root depth - Leaf Area index - Reference evapotranspiration	When water-table is shallow (wetland studies) and GW recharge is primarily influenced by evapotranspiration in root zone. Also when delay between ppt and recharge is small or not of interest.
Other means	Direct input of net recharge to the saturated zone from other means	N/A	When a sophisticated model is required. E.g. DAISY soil-plant-atmosphere model useful in agricultural studies (Abrahamsen and Hansen, 2000)

2.4.2.5. Saturated Zone (SZ) flow

MIKE SHE can simulate groundwater using the complex 3D finite difference method using the Darcy equation or a simpler linear reservoir method. The choice of which engine to use is based on the objectives of a particular study and the availability of data.

Table 2.7. Specification of Saturated Zone parameters when using the 3D finite Groundwater approach in MIKE SHE

SZ Parameter	Details	Format or Units
Lower Level	The lower level value is used to define the bottom of geological layers and geological lenses. The bottom of the geological layers is always equal to the top of the layer underneath. Values can be specified independently or relative to ground level.	.shp (polygon) or dfs2 (grid)
Upper Level	Can be defined by the lower level of the geological layer above, or the ground surface.	.shp (polygon) or dfs2 (grid)
Horizontal extent	The horizontal extent is used to define the lateral extents of geologic lenses.	.shp (polygon) or dfs2 (grid)
Horizontal hydraulic conductivity	The horizontal hydraulic conductivity is used directly in MIKE SHE. MIKE SHE assumes that the horizontal conductivity is isotropic in x and y.	$m\ s^{-1}$
Vertical hydraulic conductivity	The vertical hydraulic conductivity is input directly in MIKE SHE.	$m\ s^{-1}$
Specific yield	The specific yield is only used in transient simulations, but must always be input. Specific yield is only used in the cells that contain the water table. In the cells below the water table, the specific storage coefficient is used.	Dimensionless
Storage coefficient	The specific storage coefficient is only used in transient simulations, but must always be input. Specific storage coefficient is only used in the SZ. In the cells above the water table, the specific yield is used.	1/m
Initial potential head	The Initial potential head is the starting head for transient simulations and the initial guess for steady-state simulations. The choice of initial head for steady state simulations may affect the rate of numerical convergence depending on the solver used.	m (values relative to ground level)
Outer boundary conditions	The outer boundary conditions are defined as line segments between two boundary points.	Fixed head, zero flux, flux or gradient
Inner boundary conditions	For locations of various internal boundary conditions to be specified in the model.	Fixed head, fixed head drain, head controlled flux or inactive cells
Drainage	Surface drainage is a boundary condition in MIKE SHE used to defined natural and artificial drainage systems that cannot be defined in the MIKE 11 River setup. It can also be used to simulate overland flow in a simple lumped conceptual approach.	Drainage downhill based on adjacent drain levels, routing based on grid codes or distributed drainage.

The geology of the catchment when using the 3D finite difference method can be described with geological layers and additional geological lenses each with associated hydraulic properties. The hydraulic properties (as listed in Table 2.7) can be specified on a cell-by-cell basis or by zones defined by ArcGIS shapefiles (polygons) or MIKE SHE grid-code files. Boundary conditions also need to be specified for each computational layer.

For studies concerned with groundwater movement the finite difference method has been shown to work at a satisfactory level (Henriksen et al., 2003). In this research, groundwater flow simulation was undertaken at a large scale covering the 7330km^2 Isle of Sjaelland, Denmark, using a 1km × 1km grid. Head data from 4439 observation wells were used as the observed data for which the model was calibrated and validated. After inverse modelling was used to refine calibrated parameters, a classification of 'very good' results were obtained with RMS between 4-6, NSE 0.65 – 0.85 and F_{bal} (average runoff error) between 5-10%. Results from Henriksen et al., (2003) were considered good enough to be able to use the model for groundwater and climate change scenario testing.

However, it is frequently the case that there are problems of obtaining detailed information and data relating to groundwater, or it is not necessary to model groundwater flow in such a complex manner. In this case the conceptual linear reservoir method is frequently used as a compromise. This method, shown in Figure 2.11, divides the groundwater of the catchment into sub-catchments with a series of shallow reservoirs and one or more deep baseflow reservoirs. To account for differences in fast and slow baseflow, the baseflow reservoirs are further divided into two parallel reservoirs. The lowest interflow reservoir acts as the discharge outlet for groundwater to be added as lateral flow into the channel.

The linear reservoir approach is an example of one of the more recent developments of the model code. Studies have chosen to use this method for differing reasons, such as Andersen et al., (2001) who used this approach because of lack of saturated zone data over a large area in the Senegal Basin, and Sahoo et al., (2006) because the emphasis was on flooding and not a detailed groundwater model.

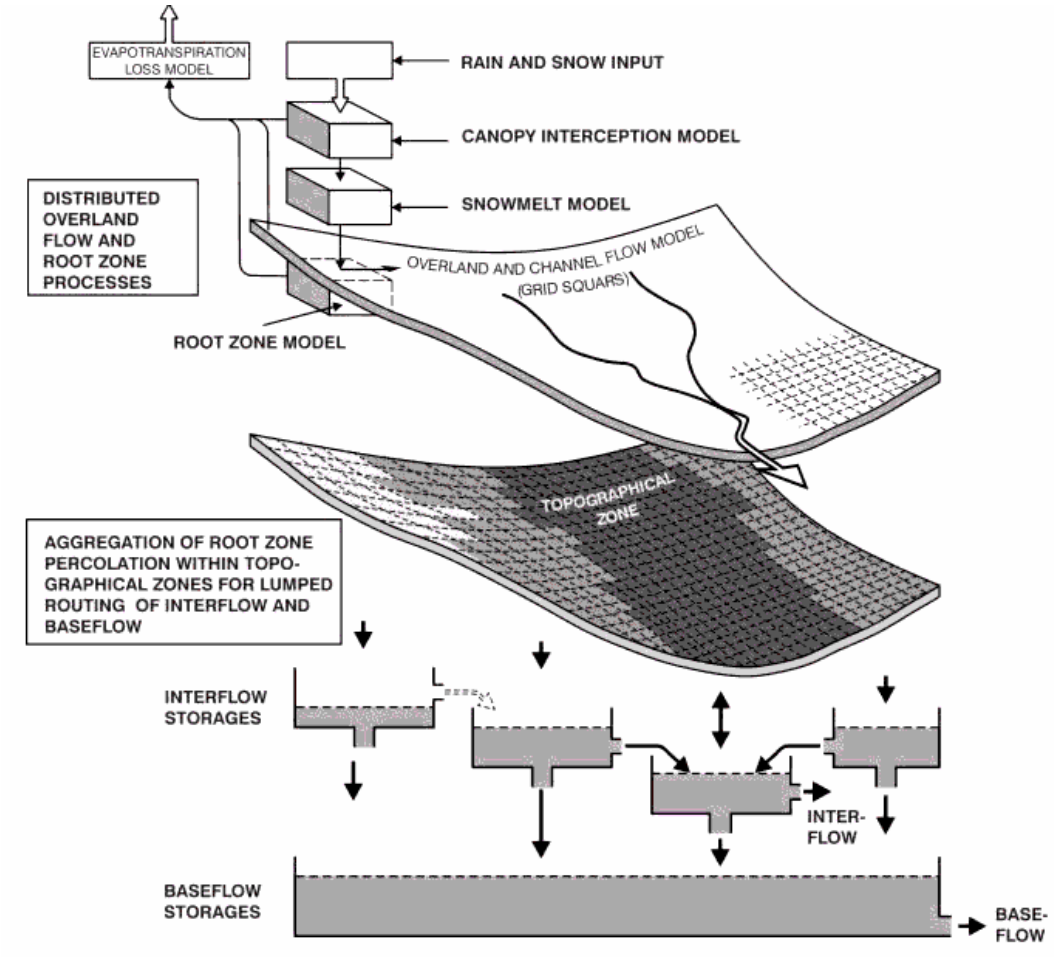


Figure 2.11.Conceptual structure of the linear reservoir approach for the saturated zone in MIKE SHE

2.4.2.6. Channel flow

Within MIKE SHE, water can enter and be exchanged within the channel (small rivulets, streams and rivers) either from overland flow or from groundwater where the channel is a discharge point based on its close proximity to the water table. This exchange is carried out with the coupling of MIKE 11, a 1D hydraulic model, based on digitized points (chainage locations) and calculation nodes (cross-section points).

MIKE SHE represents the river network along the edges of the specified model domain grid (Figure 2.12). Since the water exchange occurs along the edges of the grid it is important to note that the more highly resolved the grid, the more accurate the representation of the river network will be. Locations of MIKE SHE river links are

determined automatically with reference to the co-ordinates of the MIKE 11 river points that define the branches of the hydraulic model.

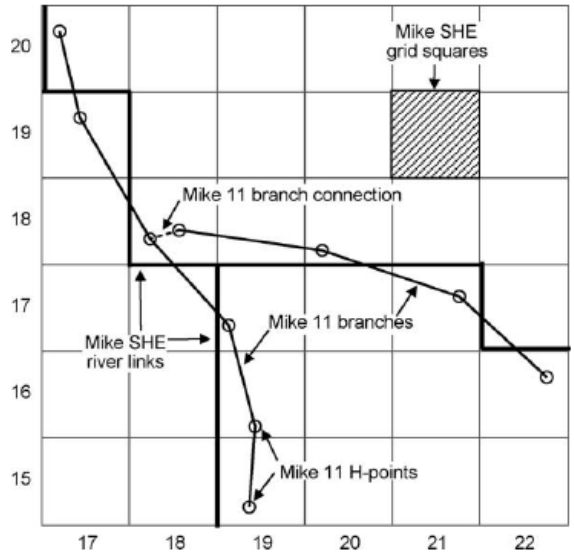


Figure 2.12. MIKE SHE and MIKE 11 coupling (DHI-WE, 2005)

As indicated in Figure 2.10 there are two methods of simulating channel flow within MIKE 11: the 1D finite difference and a more simple flow routing method. Within each method there are further options regarding the complexity of the calculations. For example, advanced and complete non-linear equations of St Venant open channel flow can be used in studies which focus on specific flow dynamics. Alternatively simple hydraulic instantaneous flow routing methods can be used which assumes a water pulse flows through the system in a single time step.

2.4.3. Applications of MIKE SHE

This section describes some of the uses and applications of MIKE SHE as collated from available literature. It is divided into two sections, the experimental research of MIKE SHE and the more operational and applied studies. The experimental and model testing research is further divided into three sections; a) general development and model comparisons, b) assessments of input data uncertainties and lastly c) research undertaken that addresses calibration and validation issues. These research studies are discussed first, followed by the presentation of a representative sample of applied research from a variety of fields/subjects. The section seeks to provide a review that

MIKE SHE is both a well tested/researched model code that is continually evolving and advancing, as well as having the ability to be used for real world problems and management issues at a variety of scales.

2.4.3.1. Experimental research using MIKE SHE

Table 2.8 summarises some of the general research studies as examples of work that have been undertaken using MIKE SHE that describe model code development and links with other specialist models. For example, Boegh et al., (2004) use MIKE SHE coupled with Daisy, a SVAT model for agricultural applications. A significant development of MIKE SHE is that the code is now OpenMI compliant. OpenMI is a tool that provides a standard interface, which allows models to exchange data with each other and other modelling tools on a time step by time step basis as they run (Moore, 2007).

Research has also been undertaken that compares the MIKE SHE code with other modelling codes, for example Yang et al., (2000), and Abu El Nasr et al., (2005). These studies confirm that MIKE SHE is able to produce models of equal or sometimes superior ability when compared to other codes. This Section of the thesis, 2.4, has ascertained that there are numerous ways in which to solve the different components of MIKE SHE. However, there does not appear to be a lot of published research on the intra-model uncertainty when comparing different models constructed using different protocols within MIKE SHE. As a result, this is an area of research that is to be addressed in this thesis.

A further class of uncertainty research that has been undertaken with the MIKE SHE model code assesses uncertainty in input data and the impacts of using different grid sizes (Table 2.9). Although these areas of research are not explicitly assessed in this thesis, they are still a primary cause of uncertainty in hydrological models that requires acknowledgement. Vazquez and Feyen (2003) and Vazquez and Feyen (2007) undertake uncertainty analyses on the different methods of inputting elevation data and potential evapotranspiration. In both cases, the different methods that were assessed were shown to have noticeable impacts on the results. This research highlights the importance of the decisions made by the hydrological modeller and should be acknowledged along with other types of uncertainty already documented.

Table 2.8. Review of general MIKE SHE developmental and model comparison studies

Research focus	Details	catchment	Author
MIKE SHE development & addition of components	Incorporation of physically-based macropore formulation using up-scaling methodology, then assesses pesticide leaching given inclusion of the macropore component.	Bjerge, 62.3km^2, Focuses on Frankerup region, 1.5 km^2 (Demark)	Christiaensen et al., (2004)
	Daisy (a soil-plant-atmosphere) model incorporated into MIKE SHE in agro-hydrological modelling that also incorporates remotely sensed data.	Field scale in agricultural region of Denmark.	Boegh et al., (2004)
	Formal methodology given for the calibration and validation of distributed models. The protocol outlines the specific need for rigorous validation and adequate spatial data. The Karup catchment is used as an example of the suggested protocol.	General with example of Karup, 440 km^2, (Denmark)	Refsgaard (1997)
	Details of initial development of MIKE SHE modelling software including schematic representation of water movement module.	NA	Refsgaard et al., (1995)
	Linking of Snow Model and MIKE SHE for a catchment with ~20% glacier coverage in Greenland.	Zackenberg River, 512km^2 (Greenland)	Mernild et al., (2008)
MIKE SHE compared to other codes	Compares three distributed modelling codes: MIKE SHE, TOPMODEL and the GB model. MIKE SHE likely to perform better in smaller catchment.	Seki River 703 km^2 (Japan)	Yang et al., (2000)
	Semi-distributed SWAT model compared to distributed MIKE SHE model using daily flows at basin outlet. MIKE SHE simulations marginally better, though models v. comparable. Important to address modelling objectives, application and availability of data.	Jeker, 465 km^2 (Belgium)	Abu El Nasr et al., (2005)

Other uncertainty analyses have been undertaken on the use of different grid sizes within MIKE SHE. Vazquez et al., (2002) undertook multi-resolution scaling assessments on the 600km^2 Grote and Kleine Gete catchments in Belgium, but identified that in this example differences in simulations were only minor.

Experimental research that uses input data from remote sensing compared to ground measurement have also been undertaken. For example Andersen et al., (2002a) and Stisen et al., (2008) that use Leaf Area Index (LAI) and precipitation derived from satellite data compared to more conventional methods. At the large scale (Senegal Basin, 350 000 km^2 and 82 000 km^2) satellite data for LAI were shown to improve the model simulations, an encouraging indication that remote sensing data may aid in overcoming some of the data gaps frequently associated with distributed hydrological modelling.

Table 2.9. Review of MIKE SHE input data uncertainty research

Research focus	Details	catchment	Author
Testing input data	Remotely sensed inputs of precipitation and LAI compared to conventional inputs. Improvements found when using RS LAI, but no real improvement to simulations with using RS precipitation.	3 sub-catchments of Senegal basin, 82 000 km^2 (Senegal)	Andersen et al., (2002)
	Different DEM gridding methods (all with resolution of 600m) are used to assess the output sensitivity in simulations. The MIKE SHE interpolation tool, data distributed about the centre of the gridded DEM cell and an ESRI TOPOGRID algorithm were tested and subject to individual calibrations to assess the variations to parameter values and adequacy of predictions.	Grote & Kleine Gete, 600 km^2 (Belgium)	Vazquez and Feyen (2007)
	Independent multi-calibrations of models using different inputs of PET. A range of Penman-based methods used to show parameters associated by evapotranspiration were sensitive to the different methods adopted.	Grote & Kleine Gete, 600 km^2 (Belgium)	Vazquez and Feyen (2003)
	Operational distributed modelling based on RS derived inputs of the most important driving variables: precipitation, PET and LAI. Quantitative analysis suggests no real improvement with RS data, but, comparison of spatial output of PET from RS derived model compared to convention point input model show significantly different spatial patterns. Research suggests RS data shows great potential for larger-scale distributed modelling.	Senegal basin, 350 000 km^2, (Senegal)	Stisen et al., (2008)
Grid size	Investigation of scale effects on simulations. Multi-resolution validation tests undertaken at 300m and 1200m grid from the calibrated 600m model to show simulations differ only marginally. A further multi-calibration test undertaken where individual model calibrations undertaken at each grid size.	Grote & Kleine Gete, 600 km^2 (Belgium)	Vazquez et al., (2002)
Uncertainty analysis	Model parameters assessed by use of transfer functions with no model calibration undertaken. Monte Carlo simulation used to assess how uncertainty in input data influences model results. The magnitude of uncertainty depended on the temporal and spatial scale.	Karup, 440 km^2, (Denmark)	Thorsen et al., (2001)
	Investigate uncertainty and error associated with flow predictions for a range of rainfall and fire conditions, in both calibration and validation and using a GLUE approach. Uses 7 scenarios of RS derived LAI to show predictive uncertainty between scenarios as less than +/- 10% & that RS derived LAI is appropriate for Chaparral catchments.	Jameson catchment, 34 km^2, California (USA)	McMichael et al., (2006)

Issues of calibration and parameterisation have also been addressed using the MIKE SHE model code. Table 2.10 provides a review of some of this research. Although it is shown that the majority of studies include multi-criteria methods of calibration that assess performance internally within the catchment at a number of locations, comparative research has not been undertaken that compares both manual and automatic calibration methods (which consequently is carried out within this thesis). In the cases of automatic calibration assessment, the presented studies do not compare automatic calibration methods or the influence different methods may have on model outputs. Despite this, Table 2.10 does indicate that issues of calibration and parameterisation have been assessed within MIKE SHE for a range of catchments of different scales and characters.

Table 2.10. Review of MIKE SHE research assessing issues of calibration and parameterisation

Research focus	Details	catchment	Author
Automatic calibration	Discusses framework for autocalibration methodology including parameterisation, specification of calibration criteria and choice of optimisation algorithm. Autocal package in MIKE ZERO used to show improved calibration from the Refsgaard (1997) model by finding the Pareto optimal models using SCE approach. A trade off is highlighted between ability to simulate GW and flows.	Karup, 440 km^2, (Denmark)	Madsen (2003)
	An inverse approach to calibration using a simplification of the system in order to reduce simulation times using steady-state (not transient) modelling. Aim is to use automatic calibration on the steady state model and then transfer to the transient model. Results show the estimated parameters are highly sensitive to the way the steady-state model is conceptualised.	South Jutland area, 5900 km^2, (Denmark)	Sonnenborg et al., (2003)
Parameter-isation	Use tests of prior information (derived from site specific field and lab studies) and non prior information (literature derived) of effective soil parameters. Two different parameter estimation studies SCE and GLUE used to show that more realistic parameterisations and reductions in uncertainty bounds found when using prior information.	Hillslope in sandy-loam belt (Luvisol), 400 km^2, (Belgium)	Mertens et al., (2004)
	Range of studies to assess sensitivity analyses and uncertainty analyses of soil hydraulic parameters. Studies include the use of GLUE within the MIKE SHE framework and results indicate soil parameters are successfully constrained to within 20 – 85% of original size.	Ohebach catchment, 1 km^2, (Germany)	Christiaens and Feyen, (2000; 2001; 2002a, b)
Calibration & validation issues	Lumped conceptual (NAM), distributed physically-based (MIKE SHE), Mixed (WATBAL) models rigorously tested in three catchments. Split sample, proxy-basin, modified proxy-basin and differential split sample tests showed models performed equally well with enough calibration data, though with no calibration the MIKE SHE model best.	Ngezi-South, 1090 km^2, Lundi, 254 km^2 and Ngezi-North, 1040 km^2 (Zimbabwe)	Refsgaard and Knudsen (1996)

Emphasises requirements for calibration and validation of lumped and distributed models. Uses multi-site and multi-criteria (river flow and groundwater level) approach. Internal validation with additional observations significantly poorer results. Model grid sizes also tested. A deterioration at coarser grids, especially greater than 1000m.	Karup, 440 km^2, (Denmark)	Refsgaard (1997)
Model performance – calibration, split sample validation and multi site validation using a multi-criteria approach (river flow and groundwater level). Multi-site validations found less accurate, groundwater level simulations variable in all tests.	Grote & Kleine Gete, 600 km^2 (Belgium)	Feyen et al., (2000)
Calibrations & validations using three tests (uncalibrated, one station, various stations) in large catchment using conventional data.	Senegal basin, 375 000 km^2 (Senegal)	Andersen et al., (2001)
Multi-validation (split sample and proxy basin tests) of various models set up over five years to simulate groundwater levels at a large scale. Highlights the non-linear process of modelling and need for continual model re-evaluation.	7330 km^2 Island Sjaelland (Denmark)	Henriksen et al., (2003)
A multi-criteria protocol to evaluate simulations during calibration and validation stages. Uses multi-objective performance statistics, visual and analytical approaches as well as filtering of flow components to add physical consistency to the modelling approach.	Nete, 356 km^2, (Belgium)	Vazquez et al., (2008)
Calibration and validation of MIKE SHE to simulate high temporal resolution 15 minute flows in a flashy mountainous catchment. Found that dividing the catchment into sub-catchments is useful in assessing sensitive model parameters (due to the heterogeneity in the catchment) before overall model calibration undertaken.	Manoa-Palolo stream, 24.6 km^2, (Hawaii)	Sahoo et al., (2006)

2.4.3.2. Applications of MIKE SHE

In addition to the research studies presented in Section 2.4.3.1, MIKE SHE has also been widely used to study a variety of water resource and environmental problems under diverse climatological and hydrological regimes (Refsgaard and Storm, 1995). This is demonstrated in Table 2.11. where it can be seen that the model has been applied to a large range of problems and locations from the very large semi-arid environment in the Senegal Basin (Andersen et al., 2001), to small scale studies in the temperate wet grasslands of the North-Kent marshes, UK (Thompson et al., 2004; Thompson et al., 2009).

Table 2.11. Selected applications of MIKE SHE from the literature

Application	Reference	Description
Modelling groundwater	Christiansen et al., (2004)	Pesticide leaching through macropores to groundwater, and role of macropore flow in groundwater recharge, 1.5km^2 basin, Denmark.
	Liu et al., (2007)	Modelling groundwater level response to variations in OLF and topography in a seasonally flooded area within the Tarim basin (Yingsu sub-basin), 91.6km^2, China.
	Smerdon et al., (2009)	MIKE SHE used to provide boundary conditions (OLF, AET, and recharge) for a large scale groundwater model, 130 km^2 BX Creek watershed, British Columbia. Study also incorporates further geochemical and isotopic analyses of GW and surface water.
Pollution and water quality	Brun (2002a, b)	Landfill pollution plumes, Denmark. Studies of pollution transport and biogeochemical processes.
Wetlands	Thompson et al., (2004)	Modelling of ditch-water levels in the North-Kent marshes, UK.
Soil erosion modelling	Storm et al., (1987)	Water flow and soil erosion processes in forest hydrology and watershed management.
Agricultural management	Boegh et al., (2004a, b)	Relationship between soil water balance and vegetation growth by coupling MIKE SHE and a vegetation-SVAT model (Daisy) to predict crop yield, Denmark.
Remote sensing	Andersen et al., (2002a, b), Stisen et al., (2008)	RS derived distributed precipitation and LAI data used to test for improvement in model performance in the Senegal Basin. RS used to derive dryness index
River basin management	Henriksen et al., (2003)	7330km^2 island off Denmark as part of national hydrologic 1km^2 grid model of Denmark.
	Jain et al., (1992)	Kolar sub catchment of Narmada, India.
	Ngo et al., (2007)	Operational modelling of the HoaBinh reservoir, Vietnam using automatic optimisation.
Irrigation	Jayatilaka et al., (1998)	9-ha irrigation site, Australia, to quantify the processes affecting surface drainage and groundwater levels.
Climate change	Thompson et al., (2009)	Impact of climate change on the Elmley marsh wetlands, UK, using UKIP02 data.
	Mernild et al., (2008)	Application of MIKE SHE in snow dominated/glacial region in NE Greenland, Zackenberg basin, 512 km^2. Climate change simulations with RCM and Hadcm3 for A2 and B2 scenarios to suggest flow magnitude will increase by factor 1.5 from current conditions.
	Singh et al., (2010)	Climate change impacts on the hydro-ecology of Loktak lake, India.

From the selected applications in Table 2.11 it can be seen that the model code has been very widely used in a variety of water resources fields since its introduction as SHE in 1986. More recent applications frequently couple the MIKE SHE model with other ecological models (Boegh et al., 2004), or employ data derived from remote sensing (Andersen et al., 2002a, b) and are computationally more complex and intensive. The MIKE SHE model can therefore be described as a dynamic model which is continually being improved and developed.

MIKE SHE can be applied at spatial scales ranging from a single soil profile (Mertens et al., 2004) which may be of use in agricultural studies where it may be of use to derive crop water requirements. As is indicated in Table 2.11, the model has also been applied to large regions that include a number of large catchments within the same model, such as the application of MIKE SHE in Denmark to create a national hydrologic model with a grid resolution of $1km^2$ (Henriksen et al., 2003).

2.5. Data and uncertainty in modelling

The purpose of this section is to further discuss issues of uncertainty within the modelling process. The emphasis in Section 2.3 was to discuss issues and uncertainty associated with model code, specifically issues arising from the use of distributed, physically based models. In addition to this 'structural' and 'parameter' uncertainty reviewed in relation to MIKE SHE studies in Section 2.4, there is additional 'run time error' and uncertainty (Ewen et al., 2006). 'Run time error' is associated with the forcing data used within the model code, such as precipitation and evaporation input errors and uncertainties.

Although the objectives of this specific research do not include assessment of the impacts of input data uncertainty on model results, it is important to briefly acknowledge this source of uncertainty in the modelling process. Some of the sources of uncertainties for different variables commonly included in hydrological models are summarised in Table 2.12. The list is not exhaustive but is included to provide an indication of some of the largest instrumentation issues, measurement issues and location issues.

The instrumentation uncertainty typically derives from limitations and difficulties with the recording equipment that can influence the quality of data that is subsequently collected. For example, issues of rain gauge or evaporation pan design or difficulties encountered when drilling boreholes or undertaking soil surveys. Likewise measurement issues including the subjectivity of the observer in reading equipment, identifying soil or rock types contribute a further degree of uncertainty in the recorded data (Dingman, 1994).

Location issues are also important in the data collection process. For example, a decision to locate a meteorological station or rain gauge in a certain location will influence the data. If the rain gauge is located too near to trees for example, then water droplets falling from leaves and shade cover may influence the recording and result in unrepresentative data. Other location issues include the difficulties faced in collecting data in inaccessible areas, although remotely sensed data can help overcome this issue.

2.5.1. Precipitation: measurement issues

Precipitation is the predominant input in the land phase of the hydrological cycle and is consequently one of the main input parameters into any catchment hydrological model. Regarding this, the issues of precipitation data collection and representation at the catchment scale are briefly reviewed as an example to demonstrate the uncertainty that input data can have upon hydrological model simulations.

Areal rainfall estimation or the volume of rainfall over a given catchment area requires an adequate number of measurements in order to capture the spatial variation across the area being assessed. It is important to ascertain the minimum density that can be tolerated for a realistic estimate of areal precipitation, not just to capture as many storm events that will inevitably be missed during their passage between gauges (Wiesner, 1970), but also to acquire realistic total volumes of water deposited over the given area during a storm event, an essential factor in hydrological modelling of river discharge.

Density and the accuracy of areal precipitation depend on the variability of rainfall across the catchment which differs between geographical areas. For example, Hutchinson (1970) showed that in mountainous areas or those of steeply sloping terrain, more gauges would be required to capture the spatial variability due to the known variation of rainfall with elevation. Secondly, the timescale of interest also determines the density needed when calculating areal precipitation. Rainfall intensities over shorter periods (e.g. hourly) are more spatially variable than if requiring daily totals, thus, a denser network would be beneficial at higher resolution temporal scales. When considering that the UK network consists of approximately 5000 gauges, it is calculated (although perhaps not with an even distribution) there is an average density of one gauge per 60km^2 (Ward and Robinson, 1990). Even when considering the relatively high density of the UK, it must be noted that for specific hydrological studies requiring areal rainfall information, either within the UK or elsewhere, then additional gauges

should be implemented to expand the existing network (O' Connell et al., 1977; Stephenson, 1968).

Table 2.12. An overview of input data uncertainty issues

Variable	e.g. Method	Instrumentation issues	Measurement issues	Location issues
Precipitation	Point raingauge measurement, radar or remotely sensed image analysis.	Robinson and Rodda (1969) described raingauge design. Evaporation from the gauge vessel and funnel surface or delayed delivery to the collecting vessel or tipping bucket especially in gauges designed to measure short term intensity. Condensation onto a cold funnel surface from a humid atmosphere.	Subjectivity of the observer in reading measurement. Although loggers can overcome this, the majority are still read 09:00 daily.	(Larson and Peck, 1974) reviewed literature on accuracy of precipitation measurement and suggested that rain gauge data are strongly influenced by any nearby shelter or obstacles; wind eddies if the site is too exposed, as well as splash-in from the ground surface WMO, 1994. How representative the gauge is of the immediate and wider surrounding area.
Evapo-transpiration	Evaporation pan or weather station parameters, large scale modelled data (e.g. MORECS).	Direct large scale measurements cannot be made. Instrumentation issues from attempts to measure Eto using evaporation pans and weather parameters such as solar and longwave radiation, air temperature, humidity, vapour pressure and wind.	Loggers are often used to record terms used in Eto calculations which reduces reader subjectivity.	Site of measurement should be typical of surrounding area (WMO, 1994, Larson, 1971). Similar issues to precipitation measurement of point measurement often used to represent evaporation over a large area.
Soil	Soil surveys.	One off field research.	Subjectivity of observer in identifying soils.	Sample site representativeness.
Geology	Geophysical surveys of boreholes.	Problems with drilling boreholes.	Identification issues.	Uncertainty whether borehole is truly representative of the surrounding area.
Land use	Land survey/ remote sensing.	Expense of remote sensing equipment or data. Interpretation of the remote sensing data. Issues of spatial resolution and pixel size (homogeneity).	Land surveys time constraining. May be quickly outdated.	Inaccessibility of land survey in remote or private land or at large scales (somewhat overcome with RS).
Topography	Land survey, LiDAR or Remotely Sensed (RS) image analysis.	Expense of undertaking LiDAR or RS survey. Similar issues to land use.	Resolution at which data are collected.	Inaccessibility of land survey in remote or private land or at large scales (somewhat overcome with RS).
River flow data	Flow gauge or stage discharge relationships.	Malfunction of recording device leading to missing data. Weirs and gauging stations can silt up/blocked with vegetation.	If river floods (out of banks) then stage-discharge relationships not valid. Flow gauge survey often impractical.	Often only measured at downstream or key locations. It is difficult to assess specific river flows further upstream.
Groundwater level data	Level recorder installed in borehole or one off measurement.	Malfunction of recording device leading to missing data.	Poor calibration of the water level recorder.	Uncertainty whether borehole is truly representative of the surrounding area.

There are several approaches that have been developed to estimate areal rainfall using point measurements collected from rain gauge data. As reviewed in Table 2.13, the methods vary in complexity with some providing only estimates of the spatially averaged precipitation in contrast to others which, in addition, seek to estimate spatially distributed precipitation across the catchment area.

Table 2.13. Summary of methods to estimate areal rainfall

Method	Detail	Benefits	Limitations
Arithmetic mean	Equally weights data from all gauges where the regional average becomes the arithmetic mean	Most simple of all areal ppt calculations. Computationally straightforward. Frequently and successfully used in lumped rainfall-runoff modelling.	Ideally only suited to homogenous and low relief areas with an even spread of gauges known to provide a representative sample (Sumner, 1988)
Isohyets	A weighted areal mean that considers the gauge distribution. Isohyets connect areas of equal rainfall for a given period of time	Better reflects the overall distribution of ppt (compared to areal mean). widely used method that depicts ppt variation in space. A logical approach.	Method depends on density of the network and the local topography. The majority of isohyetal maps are drawn subjectively but can useful in establishing general ppt trends
Thiessen polygons	Involves the construction of representative triangular polygons around each gauge. Sides of each polygon are constructed along the perpendicular bisector of lines joining each pair of adjacent gauges. (Thiessen, 1911).	Attempt is made to spatially distribute data. Especially useful where there is an irregular distribution of rain gauges. If there is a high density of gauges within close proximity, the area weightings are kept small so as not to over bias specific regions within the catchment.	No consideration given to variation in ppt resulting from topographical differences. Polygons can result in distinct differences along the boundaries which can be unnaturally abrupt from one polygon to the next.
Inverse distance weighting (IDW)	Deterministic surface fitting technique. Uses measured point values from gauge locations to specify a surface representing ppt at all points within the catchment	Considered a fully distributed method	Can be considered as moderate to high computational complexity.

2.5.2. Precipitation: uncertainties in hydrological model simulations

In order to exemplify that the issue of uncertainty in input precipitation in hydrological modelling is important, several studies have underlined the high degree of sensitivity of hydrological model simulations to the spatial and temporal variability of rainfall input at the catchment scale (Krajewski et al., 1991; Shah et al., 1996; Georgakakos et al., 1996; Koren et al., 1999; Sun et al., 2000; Carpenter et al., 2001; Carpenter and Georgakakos 2004a, b). As noted by Chaplot et al., (2005), the ability of any environmental model to predict outputs at the catchment outlet depends a great deal on how well the available spatial data describe the catchments characteristics. As a result, the sensitivity of

catchment models to inaccurate rainfall input has become a key issue in the field of hydrological modelling.

The listed studies have been undertaken within a range of different catchments with different characteristics (size, geology, topography). For example, a 6.73km^2 experimental research catchment in Arizona was assessed by Lopes (1996), whereas Duncan et al., (1993) assess sampling density in the large (4800 km^2) Yamaska Basin, Canada. The studies assess different methods of inputting precipitation into the different modelling codes used to undertake the research. For example Chaplot et al., (2005) and Lopes (1996) undertake research into including a range of different densities of rain gauge data and assessing the differing impacts on output river flows. In contrast, other research assesses different methods of data collection in comparing precipitation input from rain gauges compared to radar data (Duncan et al., 1993; Carpenter and Georgakakos, 2004a; b).

Although this review has been primarily specific to precipitation data uncertainty, it demonstrates that not only the quality, but the input data method (a choice made by the hydrological modeller) can be important in determining model results.

2.6. Model calibration and validation

This section reviews issues and methods of model parameterisation, calibration and validation within hydrological catchment modelling. Methods used to derive model parameter values and performance measures that are used to test model calibrations are first reviewed. An outline of the need for model calibration is then given with manual and automatic calibration methods and parameter sensitivity analysis discussed. A summary of methods of model validation and testing conclude the section.

2.6.1. Acquisition of parameter values in hydrological models

Parameter values can be obtained by direct measurement but it is more frequently the case that the values are derived from trial-and-error processes ranging in complexity from simple ‘curve-fitting’, to more technical automatic calibration approaches such as the Generalised Likelihood Uncertainty Estimation (GLUE) technique (Beven and Binley, 1992; Freer et al., 1996).

Direct field measurements of parameter values are preferred, especially when undertaking high spatial resolution studies as they use real world data for the specific catchment being studied. However, this method of deriving parameter values is not cost effective and is difficult, especially in data sparse and remote areas (Andersen et al., 2001; 2002a, b; Van der Linden and Woo, 2003a, b). A key problem associated with direct field measurement to obtain parameter values is that it does not always eradicate the error and uncertainty outlined in up-scaling of sub-grid processes to the grid scale as discussed in the previous section. The increasing capabilities of remote sensing and the direct use of GIS in hydrological modelling, as reviewed by Chapman and Thornes (2003), is proving useful and may signify a new paradigm shift – as sought after by Silberstein, (2006). Remote sensing is the only reasonable means to obtain comprehensive spatially distributed hydrological information at a reasonable cost (Biftu and Gan 2001), and as such is more frequently being used in model input and to aid in the calibration of models.

An alternative to direct field measurement is indirect acquisition of parameter values from other studies that have undertaken hydrological field measurements. There are many hydrological modelling studies using MIKE SHE that have derived parameter values from the literature such as Andersen et al., (2001) at a large scale of 350 000km^2 in the Senegal Basin as well as Thompson et al., (2004) on a small scale of 8.7km^2 in the North Kent Marshes in the UK. Henriksen et al., (2003) adopt and refine parameter values measured within the Danish geological database as part of their study on the Isle of Sjaelland. This method of parameter estimation is therefore considered as a widely adopted approach in hydrological modelling and is of use if prior research has been carried out within the study site in question or for an area with similar characteristics.

Despite this method being considered as better than ‘guessing’ the values for hydrological parameters, there are transferability limitations that need consideration. Notably, the parameter values that are adopted have not always been derived for the specific site, and may not be representative as has been shown when a calibrated parameter values for a particular basin are adopted for distant or even neighbouring catchments (Van der Linden and Woo, 2003b). However, the adoption of ‘prior information’ (whether from the literature or laboratory research undertaken from field samples) have been shown to improve the estimation of effective parameter values in

hydrological modelling (Mertens et al., 2004), in comparison to using no prior information at all.

2.6.2. Measures of model performance

In order to quantifiably assess hydrological model simulation performance in comparison to measured data during the calibration and testing phases, there have been many statistical tests that have been developed. Performance tests usually involve comparing observed and simulated data such as groundwater levels and river flow at the basin outlet. Within MIKE SHE, in-built calculations of statistics are included the completion of each model simulation (DHI-WE, 2005). The Root Mean Square Error (RMSE), correlation coefficient (R), and Nash-Sutcliffe coefficient (NSE) are some of the key statistics assessed within MIKE SHE (Table 2.14). A brief review of each of these statistics are given in the table, with it being shown that the statistics assess different aspects of the simulation. The RMSE provides a measure of the average magnitude of error and the correlation coefficient, R, the general 'goodness of fit' (Legates and McCabe, 1999). The Nash-Sutcliffe NSE (Nash and Sutcliffe, 1970) was developed specifically for the purpose of hydrological modelling and as identified by McCuen et al., (2006) has been a widely used and potentially reliable statistic that also measures the goodness of fit.

Research has been undertaken that comprehensively reviews many of the performance statistics (Legates and McCabe, 1999; Krause et al., 2005) but in the majority of studies where hydrological models are developed and applied, usually only three or four performance statistics are assessed. For example, Zhang et al., (2008) constructed a MIKE SHE model within the loess plateau region in China. The model performance was assessed using the correlation coefficient, R, the coefficient of determination, CD (also listed in Table 2.14), the RMSE and F_{bal} which is a percentage measure of overall mean stream flow error. Sahoo et al., (2006) also used the MIKE SHE code in a mountainous catchment in Hawaii. Model performance was assessed using three statistical measures; R, RMSE and the mean error. Research undertaken using other modelling codes also assess models with similar methods, for example Abu El Nasr et al., (2005) that compared model performance of both MIKE SHE and the SWAT model.

In addition to these quantitative statistics that describe model performance, qualitative model performance is also frequently assessed in visually comparing plots of observed and simulated data (Jayatilaka et al., 1998). This research seeks to include a thorough assessment of both quantitative and qualitative methods. In this research, the development of the 'summary score' measure is described in Chapter 5 which seeks to summarise many of the qualitative indicators of performance and attribute a numerical scoring system. As suggested by Refsgaard (1997) and Refsgaard (2000), the performance of the models will also be assessed using a multi-criteria approach at internal river flow gauging stations and also a number of groundwater level boreholes.

Table 2.14. Statistical tests: RMSE, NSE and R

Statistic	Description
Root mean squared error (RMSE)	The RMSE is a quadratic scoring rule which measures the average magnitude of error. Expressing the formula in words, the difference between forecast and corresponding observed values are each squared and then averaged over the sample. The square root of the average is then calculated. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE is most useful when large errors are particularly undesirable. RMSE can range from 0 to ∞. They are negatively-oriented scores with lower values indicating a better performance.
Pearson product moment Correlation Coefficient (R)	A measure of how well the predicted values from a forecast model "fit" with the real-life data. The correlation coefficient is a number between -1 and 1. If there is no relationship between the predicted values and the actual values the correlation coefficient is 0 or very low (where the predicted values are no better than random numbers). As the strength of the relationship between the predicted values and actual values increases so does the correlation coefficient with a perfect fit giving a coefficient of 1. The linear correlation coefficient is also referred to as the 'goodness of fit' measure (Legates and McCabe, 1999) or Pearson product moment correlation coefficient. The statistic is also related to the frequently used r^2 coefficient of determination which is the ratio of the explained variation to the total variation and scored between 0-1.
Nash-Sutcliffe coefficient of efficiency (NSE)	The coefficient of efficiency E proposed by Nash and Sutcliffe (1970) is identified as one minus the sum of the absolute squared differences between the predicted and observed values normalized by the variance of the observed values during the period under investigation (Krause et al., 2005) The result ranges from -∞ to 1. An efficiency of 1 (E=1) corresponds to a perfect match of modelled discharge to the observed data. An efficiency of 0 (E=0) indicates that the model predictions are as accurate as the mean of the observed data, whereas an efficiency less than zero ($-\infty<E<0$) occurs when the observed mean is a better predictor than the model. The closer the model efficiency is to 1, the more accurate the model is.

2.6.3. Calibration

Model calibration is an integral part of hydrological modelling in which values for model parameters are adjusted to enable the model to closely match the behaviour of the system it represents (Kirkby et al., 1993 Gupta et al., 1998).

There is an unavoidable need for calibration especially in distributed modelling as the many parameters in a model are likely to comprise various values for different land covers or soil types for each computational grid cell (Bingeman et al., 2006; Madsen, 2003), especially where the specified grid size in a model is large. In addition, as all the components of the hydrological system can be included within models such as MIKE SHE, there are substantially more parameters that require calibration compared to conceptual, lumped models. This raises an issue of uncertainty within the parameterisation since it is possible that physically-based models appear to simulate the 'correct' river flow yet in reality model the hydrological processes incorrectly (Beven 1989).

Overparameterisation is a significant problem with complex hydrological models (Beven, 1996; Refsgaard, 1997; Perrin et al., 2001). Refsgaard (1997) recommends keeping the number of free parameters that require calibration at a reasonably low level in order to reduce overparameterisation. It is expected that the more field data input to the model, the more physically realistic the model parameters become, however, there is a trade off in that this also increases the number of parameters that may need calibration.

2.6.3.1. Manual calibration

Methods used in manual calibration of hydrological models have evolved with the development from lumped, conceptual models to more complex physically based and distributed models. The classical approach, used primarily by lumped, conceptual models is to undertake calibration by 'curve fitting' to observed river flow data. Curve fitting is concerned with the adjustment of the calibrated parameters until the best-fit with the observed data can be found, as indicated by model performance statistics.

A key issue of the curve fitting method is that it requires a long time series of data. This is frequently a limiting constraint as in many areas adequate records do not often exist. In addition, curve fitting results in any parameter values losing any 'real-world' meaning they may have had since the process of curve-fitting is not physically-based (Abbott et al., 1986a). Attempts have been made to refine the technique of curve fitting by separating the base flow and surface runoff components such as work by Abu El-Nasr et al., (2005).

In addition, curve-fitting techniques are not suitable for physically based, distributed models as they often have many more parameters than lumped or conceptual models that need calibrating. This is a problem since this method of calibration seeks to find an optimal parameter set thought to be representative of the catchment area (Freer et al., 1996). It has been shown in the literature, however, that there may be many acceptable parameter sets within a model that can simulate river flow for a catchment – but these can often come from different regions of the parameter space, an issue termed 'equifinality' (Freer et al., 1996). Equifinality is a result of error and uncertainty in representation of the hydrological processes and boundary conditions (Stephenson and Freeze, 1974), within the modelling process. Techniques such as Generalised Likelihood Uncertainty Analysis (GLUE) and more recent methods of automatic calibration that are used to quantify and assess the equifinality problem are detailed further in Section 2.6.3.3.

2.6.3.2. Sensitivity analysis

The process of parameter sensitivity analysis can be regarded as a tool that is useful in highlighting key model parameters that are important in influencing the model results when modified or perturbed. As defined by Sieber and Uhlenbrook (2005) parameter sensitivity can be used to describe the influence of the values of a parameter on the simulation results. When changes to the defined parameter values result in large influences on the simulated results then the parameter is classed as sensitive. The process is a key step within the hydrological modelling protocol as it can help indicate whether the models reactions to parameter perturbations are realistic, therefore confirming basic operational ability of the model and improving confidence in the model structure.

The issue of parameter sensitivity analysis includes a wide field of research (e.g. Hamby et al., 1994) where relatively simplistic and computationally complex methods are described. Manual methods of parameter sensitivity analysis were successfully undertaken by Xevi et al., (1997) in the 1km^2 Neunkirchen research catchment, Germany, using MIKE SHE. They used uniform parameter values for the catchment and modified them one by one. The results were assessed by inspection of a range of indices such as hydrograph peak and cumulative outflow at the catchment outlet.

Similarly, Brath et al., (2004) used manual parameter sensitivity analysis within a distributed rainfall-runoff model in the 1050 km^2 Reno catchment in the Apennine Mountains, Italy. The method involved specifying parameter values of + and – 10% of the feasible range for each parameter and comparison of flows with the non-perturbed set.

(Crick et al., 1987 cited in Hamby, 1994) termed this method of sensitivity analysis as 'local' since it only addresses sensitivity relative to the point estimates chosen and not for the entire parameter distribution. Despite this it is a popular method as it is technically easy when compared to more complex methods and produces useful results.

Two methods of parameter sensitivity analysis were compared by Sieber and Uhlenbrook (2005) for 40 km^2 and 15.2 km^2 basins in the Black Forest Mountains, Germany using the Distributed Tracer Aided Catchment (TAC) model. A regional sensitivity analysis (RSA) also known as generalised sensitivity analysis or the Hornberger-Spear-Young method (Spear and Hornberger, 1980, cited in Sieber and Uhlenbrook., 2005) allowed eight significant parameters to be ascertained using a graphical method. A more complex regression-based sensitivity analysis that considers time-dependency as a factor in the parameter sensitivity of a model was also used. It was successfully shown that the sensitivity of models parameters was time dependent and not uniform throughout the model time-period. However the method was also noted as time consuming with only few other studies that have successfully used it.

Parameter sensitivity analysis is described as a necessary and beneficial process. Although complex and technical methods of sensitivity analysis exist, the key issue is that the process (even if undertaken using a relatively simple local methods) should be carried out to improve the calibration and confidence in the model ability (Saltelli, 2000)

2.6.3.3. Automatic calibration

As a result of the equifinality problem identified in Section 2.6.3.1 there has been a need for a new paradigm in model calibration – the development of a calibration methodology that is more suited to physically-based, distributed models.
Generalised Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) is a Bayesian Monte Carlo simulation-based technique which was developed from Generalised Sensitivity Analysis (GSA) (Spear and Hornberger, 1980). The GLUE framework was developed as a result of the recognised 'equifinality' of models and parameter sets. It is suggested that if there is no unique optimal model, then parameter sets should be considered on a scale of likelihood. It is important to note that one parameter set may yield the best results although it is likely there will be many other parameter sets which simulate observations to a similar level. In addition, when further observational data are considered, then the rank of the best parameter sets is likely to have changed.

Implementing GLUE requires the use of Monte Carlo simulation to derive a large number of parameter sets. The relative performance of each set is assessed by comparing model estimates with observed data, and retaining only those parameter sets that yield acceptable results. The methodology has been successfully used by Freer et al., (1996) in a small research catchment in the Vosges, France using TOPMODEL. McMichael et al., (2006) used GLUE to study uncertainty in remote sensing based LAI estimates for semi-arid shrublands in California using MIKE SHE. Christiaens and Feyen (2002a) also used MIKE SHE applying a GLUE framework to constrain the soil hydraulic parameter for the $1km^2$ Ohebach catchment in Germany. Such studies have shown GLUE as a useful tool in addressing the equifinality issue.

Automatic calibration such as that available within the MIKE ZERO modelling interface of which MIKE SHE is a part, involves Monte-Carlo sampling of values of specified parameters within given parameter ranges. Optimisation of the specific parameters are found by repetitions of model runs with different values (such as in the GLUE methodology) until a given stopping criterion is met. Alternatively, a solution may not have been found after the specified number of model evaluations, in which case refinement and repetition of the process needs to be undertaken (Madsen, 2003). As a guide, the key methodology of automatic calibration follows these given steps:

- Model parameterisation and choice of calibration parameters – A sensitivity analysis can be undertaken in which parameters are successively varied one at a time. In this way an assessment can be made as to which parameters affect model performance the most, and consequently these may be selected as the parameters which are subject to automatic calibration.

- Specification of calibration criteria – This is concerned with the definition of objective functions. The objective functions seek to compare observations with simulated values, usually within a multi-objective framework of multi-variable and multi-site measurements (Madsen, 2003; DHI-WE, 2005). The calibration is then solved using a single objective function that aggregates all the specified functions. These can be weighted to reflect the importance of certain objectives of the research.

- Choice of optimisation algorithm - in MIKE SHE there are two optimisation methods that can be selected. The Shuffled Complex Evolution method (SCE) and the Population Simplex Evolution (PSE) which are both global optimisation algorithms suited to non-linear systems such as integrated catchment modelling (DHI-WE, 2005). These algorithms simultaneously evolve numerous potential solutions towards the region of the global optimum of the objective function as reviewed by Madsen (2003).

The widely used SCE is an optimisation method which combines the strengths of previous techniques to give a robust method for the automatic calibration procedure (Duan et al., 1993). It includes aspects from the simplex procedure (Nelder and Mead, 1965) with the added concept of controlled random search (Price, 1983), which is based on systematic evolution of a complex of points within the parameter space in the direction of global improvement (Holland, 1975). The process is based on a feedback loop to refine and improve parameter results until given criteria or test statistics are met.

Madsen (2003) used the inbuilt automatic calibration tool ‘autocal’ within the MIKE ZERO package for the Karup catchment, Denmark. Improvements were found in runoff simulation when autocal was used compared to the expert manual calibration from Refsgaard (1997). Groundwater level simulation however did not show any significant improvement when autocalibration was used.

An additional study by Mertens et al., (2004) used the automatic calibration tool within MIKE SHE on a hill slope for nine soil moisture measurement locations in the sandy-loam belt of Belgium. The research positively identified the importance of prior information to be incorporated within the SCE parameter estimation methodology. Prior information included estimates of the soil parameters obtained from laboratory measurements on samples collected from the field for the 11 effective parameter values within the model. The effective parameter combinations were found to be more realistic with prior information of the parameter space included. This study additionally highlighted the use of a combination of field techniques to inform the automatic calibration procedure.

2.6.4. Validation

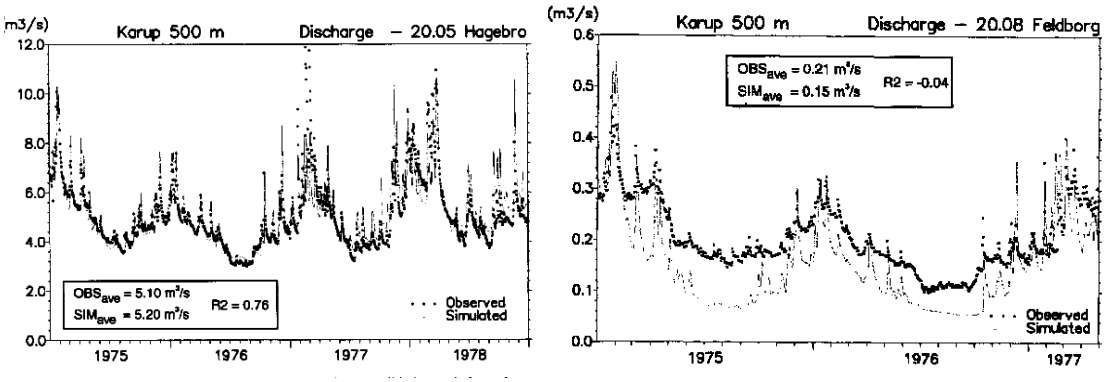
Once a model has been calibrated, its usefulness and reproducibility outside of the time period or calibration site it has been set for must be evaluated (Klemes, 1986; Tsang, 1991). Model validation seeks to assess whether the model possesses a satisfactory range of accuracy consistent within its intended application (Schlesinger et al., 1979).

2.6.4.1. Split sample traditional calibration and validation

Although there has been no agreement on a uniform methodology of model validation (Vogel and Sankarasubramanian, 2003), the work of Klemes (1986) has proved fundamental in the majority of hydrological modelling studies. The most widely adopted framework for model calibration and validation is the split sample approach in which the time series of data are split (ideally into equal time lengths), with one being used to calibrate the model and the other used to run the simulation with the calibrated parameters as a model validation. It is suggested that the model can be deemed acceptable if the two simulations give statistically similar results within the allowed error margin (Klemes, 1986).

As with the classical curve-fitting calibration technique, the traditional simple split sample of time-series validation methodology is not always adequate for complex models because of the increase in the number of model parameters. Refsgaard (1997) describes how the number of outputs from spatially distributed and physically based models are numerous (e.g. river flow, groundwater level, soil moisture) and therefore

simply comparing observed with simulated outflow at the basin outlet for two separate time periods does not fully test all the components within the model. In the study the traditional split sample test resulted in adequate model performance. It is highlighted in Figure 2.13 that the NSE model performance statistic decreases from 0.76 at a site used in the calibration to -0.04 at a site not used in the calibration.



(a) (b)

Figure 2.13. Results of multi-criteria validation (Refsgaard, 1997) illustrating the need for the incorporation of a rigorous methodology in the validation of distributed models. a) displays the results of the calibration at the catchment outlet, and b) the validation at another internal gauging station for which there was observed data.

2.6.4.2. Multi-criteria comprehensive validation

Klemes (1986; 1983) discuss the possibility of calibrating catchment hydrological models against a number of spatially distributed internal state variables in multi-criteria evaluation, not just river flow at the outlet of the catchment. Abu El-Nasr et al., (2005) employ this method of model evaluation where multi-site validation within MIKE SHE was used to test the model using river flow data as well as observed time series of groundwater levels. Additionally, Thompson et al., (2004) undertake a similar multi-site study in which different channel water level gauge sites are used for the calibration and validation stages within a wetland context. Refsgaard and Knudsen (1996) undertook a thorough analysis comparing and validating three hydrological models in three different sized catchments in Zimbabwe. The study is likely the most rigorous in the literature that tests the four basic categories of typical modelling tests (Klemes, 1986) as reviewed in Table 2.14, including the split-sample test (SS), the differential split-sample test (DSS), the proxy-basin test (PB) and the proxy-basin differential split-sample test (PB-DSS).

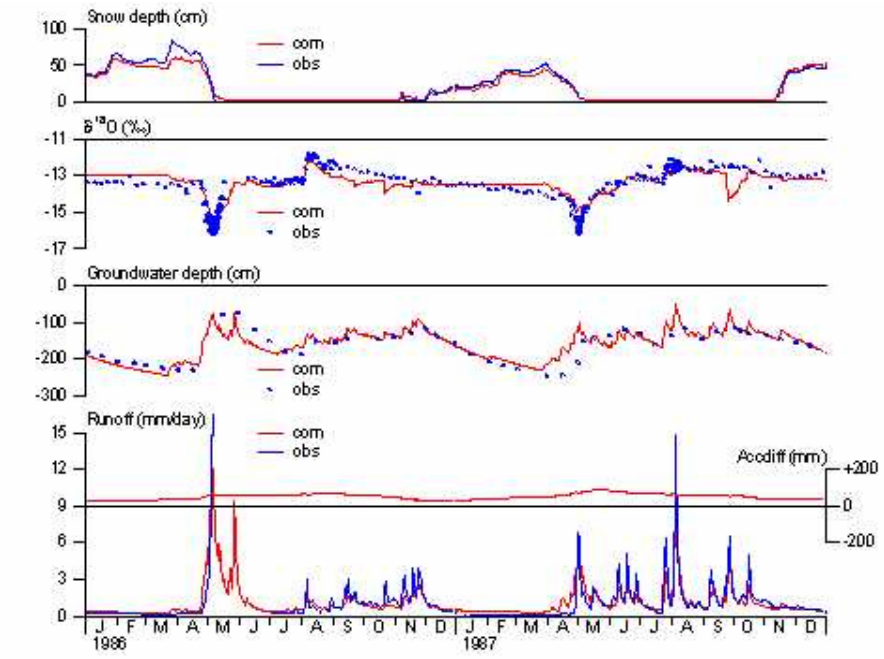


Figure 2.14. Integrated multi-objective calibration and validation, a case study of the Swedish meteorological Hydrological Institute
Source: http://www.smhi.se/sgn0106/if/hydrologi/projects/proj_r11.htm

Table 2.15. Review of operational testing of hydrological simulation models (Klemes, 1986)

Test	**When to use**	**Method**
Split-Sample (SS)	Testing river flow in a gauged basin with adequate time series of data	When sufficiently long, the record is split in two equal parts – one for calibration, the other validation. Otherwise, the record should be split 70% calibration, 30% validation. The model deemed acceptable if both calibration and validation results are similar and errors minimal.
Differential Split-Sample (DSS)	When model is to simulate flows in a gauged basin under conditions different from those corresponding to the available flow record (e.g. change in climate)	Two periods with the different climatic parameter of interest selected from the available record (e.g. high and low mean precipitation). If model intended to simulate wet climate flow then it must be calibrated on dry record and validated with the wet record (vice versa). In a case of land use change the DSS must be carried out on two substitute basins (e.g. with both models calibrate model prior to land use, validate model on the changed land use to assess ability of model to simulate flow given the transition in question).
Proxy Basin (PB)	A basic test for geographic transposability to a separate basin	For un-gauged basin C, gauged basins A and B are selected within the same region. Calibration undertaken on basin A, Validation on basin B. The model is considered sufficient and transferable to C if results are similar and there are minimal errors for A and B.
Proxy Basin, Differential Split-Sample (PB-DSS)	Where the model is supposed both geographically and climatically (or land use wise) transposable	Different forms depending on specific modelling task. E.g. for assessing flow under climatic change in un-gauged basin C:2 gauged basins, A and B with characteristics of those of C identified with Wet and Dry periods for each selected. (Aw, Ad and Bw, Bd). To assess a wet climate in basin C, Ad/Bw and Bd/Aw need to be undertaken for calibration/validation respectively. The model is judged adequate if results from Bw and Aw similar.

Figure 2.14 reviews how multi-objective calibration and validation for a number of observed parameters is used for the HBV model for the 0.5km^2 Svartberget research catchment, Sweden. The Swedish meteorological hydrological institute (SMHI) used snow depth, oxygen 18 concentrations, groundwater levels and river flow to test the model, where a good calibration is shown for all parameters. The ability of a model to simulate and be tested with numerous parameters improves the confidence in results, that the model is simulating a range of processes well, rather than just the ability to simulate one parameter such as river flows. Such methods of testing models are however difficult to undertake as they require a vast body of observed data for a range of different parameters.

The issue of temporal and spatial transferability of calibrated parameters and the applicability outside of the calibrated set-up is another issue of debate and uncertainty in the literature that has not yet been clearly established. Unrealistic results may occur whenever previously calibrated parameters are applied to another model as different physical processes, and their associated parameters may become more important during different periods, or under changed environmental settings (Apaydin et al., 2006). Van der Linden and Woo (2003b) studied the transferability of parameter estimates within the 277,100 km^2 Liard Basin, western Canada. Parameter values were transferred both across catchments with close proximity, and internally within the study catchments' sub-basins of differing sizes. Parameter values that were calibrated internally for each of the sub-basins (dashed lines) (Fort Nelson and Kachika basins, ~10,000 km^2) indicated a better performance than values derived for parameter values calibrated at the basin outlet (solid line), as identified in Figure 2.15.

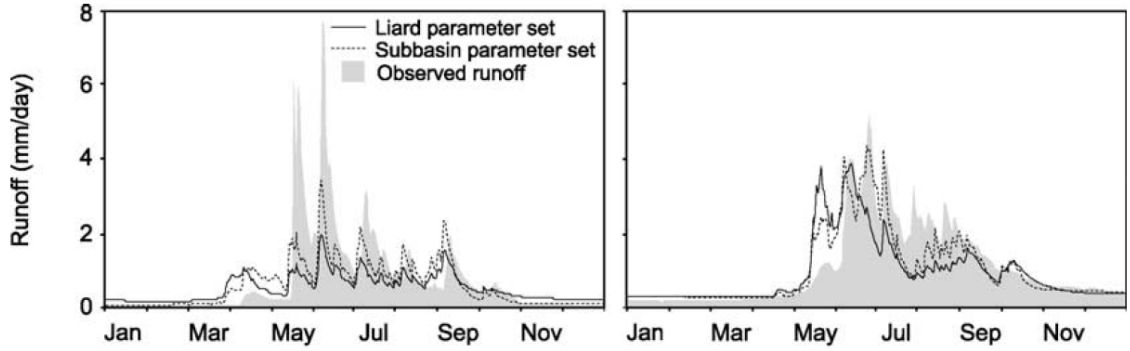


Figure 2.15. River flow calculated with parameter calibration for specific sub-basins compared to Liard (basin outlet). Shown for two internal gauging stations a) Fort Nelson sub-basin and b) Kachika basin. Source. Van der Linden and Woo (2003b)

The results aid in demonstrating the problems of transferability of parameter values. In a comparable study, Heuvelmans et al., (2004) concluded with similar findings to Van der Linden and Woo (2003b). Transferability of calibrated values were investigated at the catchment scale, a neighbouring catchment, and a catchment with a different environmental setting. Decline in model performance was found when model parameters were transferred – especially in the basin with different characteristics.

When calibrating and validating hydrological models there is always a possibility for improvement. The more parameters that need adjusting through calibration, the more field data are required and the more work is required by the modeller. As Klemes (1986) outlines, a model should be validated according to the type of application for which it is intended. It is often the case that hydrological modelling is not a linear process but as detailed in Henriksen et al., (2003) can include many feedback loops in which continual model re-evaluation is necessary.

2.7. Summary

This chapter has introduced the principal concepts, issues and uncertainties associated with hydrological modelling, as well as introducing and discussing the MIKE SHE model code that is subsequently used in this research. Given the review it can be seen there are a large range of issues of uncertainty that it is possible to assess within the hydrological modelling process. Although research has previously been carried out into hydrological model uncertainty, there has been an identified gap in the literature in comparing hydrological models that have been developed using the same model code but with different protocols (the method of spatial distribution with reference to different calibration methods), an area of research that this thesis seeks to address.

MIKE SHE was shown as a suitable code in which to assess issues of uncertainty as it has been described that it is possible to construct models at differing spatial complexities, with different parameters (objectives of the research), whilst at the same time being able to construct the models within the same framework.

Chapter 3 describes the Tern catchment that is used as a case study in this research. A comprehensive catchment review and data analyses are undertaken for the data that are to be used in the construction of the MIKE SHE models in Chapter 4. Chapter 3 provides the conceptual understanding of the catchment that is required before hydrological models are constructed.

Chapter 3

The Tern Catchment: Synthesis of hydrometeorological and related data

3.1. Introduction

The Tern catchment, an upstream tributary of the River Severn, covers an area of 876.36km^2 centred over the North Shropshire plain, UK, and extends between SJ 340630 and SJ 385634 on the British National Grid. As highlighted in Figure 3.1, the region is drained by the River Tern and its main tributaries, the Meese, Roden and Strine. The catchments of these rivers are rural but include a few small towns and villages such as Market Drayton, Wem and Newport. The larger urban area of Telford partly lies within the catchment, and is located to the south.

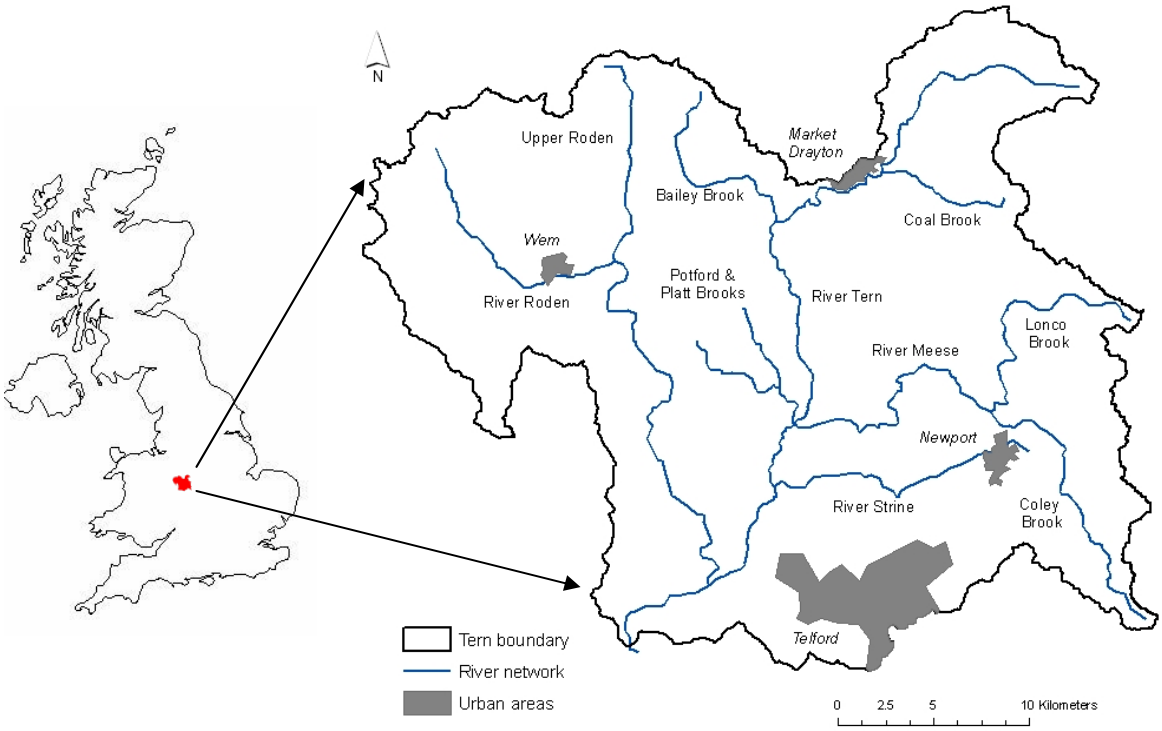


Figure 3.1. Location of the River Tern Catchment

This chapter describes the Tern catchment and the data used within this thesis. Firstly, a brief summary of the catchment is given, followed by detailed descriptions and analyses of separate components of the system that affect it's hydrology, these include; basin topography (Section 3.2), the geological setting including details of solid geology and drift deposits as well as a hydro-geological analysis (Section 3.3). Also provided is a review of catchment soils, for both the England and Wales soil associations and hydrology of soil types (HOST) classification (Section 3.4). The distribution and characteristics of different land cover classes are given in Section 3.5.

The chapter includes both a review of these secondary data as well as new analyses undertaken to provide a comprehensive review of the catchment. Section 3.6 provides analysis of catchment meteorology including a detailed characterisation of long-term catchment precipitation (1975-2000) as well as a further analysis of precipitation for 1999-2003, the modelling period used in Chapters 5, 6 and 7. Section 3.6.2 describes and compares the calculation of Hargreaves-Samani evapotranspiration data with other UK based evapotranspiration methods. A detailed analysis of river flow is provided in Section 3.7 including a description of the river network, statistical analyses of flow at different gauges, as well as qualitative reviews of the gauges to aid in quality control and understanding of the flow data. A review is also provided of the Shropshire Groundwater Scheme (SGS), a system designed to augment downstream river flow during dry periods. The penultimate Section, Section 3.8, details groundwater levels recorded from a series of different boreholes in differing geologies. Section 3.9 provides an overview of the water balance for the catchment.

3.2. Topography

High spatial resolution Ordnance Survey Land-Form PROFILE DTM 1:10,000 Topographic data for the Tern catchment have been freely acquired from the Edina Digimap service (edina.ac.uk/digimap) and pre-processed to an ASCII grid format. The Digital Terrain Model (DTM) data have a 10m horizontal grid interval, 1m height resolution and have a height accuracy of 2.5 - 5m (edina.ac.uk/digimap/ description/products).

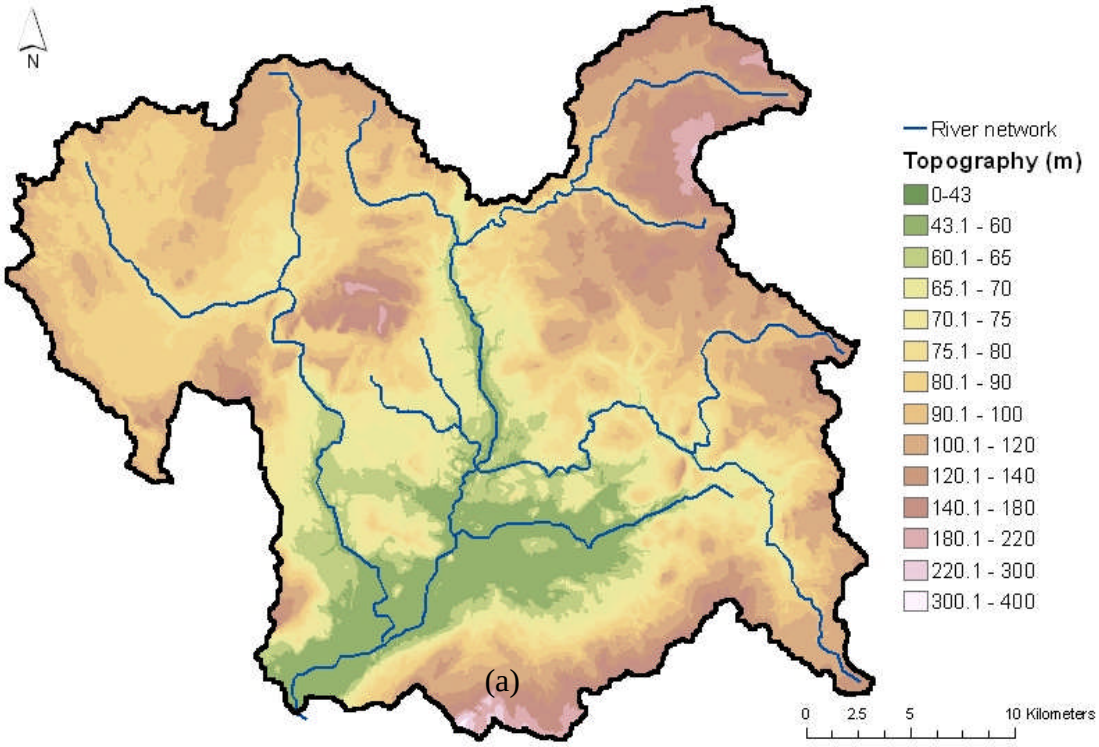


Figure 3.2. Topography and stream network of the River Tern Catchment
(a) denotes the location of the Wrekin, the highest point in the catchment

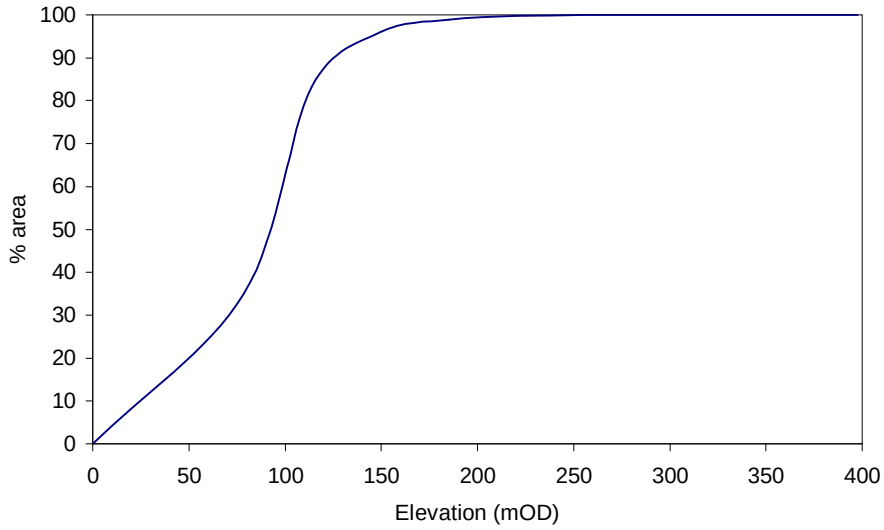


Figure 3.3. Hypsometric curve for the Tern Catchment

Table 3.1. Topographical catchment characteristics of the Tern Catchment

Feature	**Value**
Catchment area	876.36km^2
Maximum elevation	397.7 m
Minimum elevation	42.9 m
Mean elevation	91.3 m

Figure 3.2 shows the topography and river network within the catchment is predominantly low lying with gentle undulations along the River Tern which rise to escarpments on the hills, especially to the north-east. In the south-central extremity, the north slope of the Wrekin (397.7 mOD) is the highest point in the catchment; this is located just east of the basin outlet and point of minimum elevation at 42.9 mOD. The mean elevation within the catchment is 91.3 mOD and the area-elevation curve (Figure 3.3) highlights the basins low lying characteristic, with 80% of the catchment being below 120 mOD. Summary elevation statistics are provided in Table 3.1.

3.3. Geology

The spatial extent and rock types within the Tern catchment have been available from the British Geological Society (BGS) for research purposes, as a result of the LOCAR project (Section 1.4). The data are also now available from the geology section of Edina Digimap. Additional information on the hydrogeology of the basin is amalgamated within Adams et al., (2003), a report of the LOCAR hydro-geological infrastructure for the Tern catchment.

3.3.1. Geological setting and history

The Tern catchment is divided between two geological basins – the southern margin of the Cheshire Basin to the Northwest, and Stafford Basin to the east. The dividing line between the two basins is the Hodnet fault (highlighted in Figure 3.4) that runs north-east to south-west.

The underlying bedrock within the Tern catchment is predominantly comprised of permeable Permo-Triassic Sherwood sandstone (Figures 3.6 and 3.6) with overlying drift deposits that influence groundwater recharge in the basin. The aquifer thickness varies over short distances (between 50-200m) as a result of the complex fractures and fault lines across the catchment (Adams et al., 2003). The major fault lines are highlighted in Figure 3.4, which predominantly lay in a SW to NE direction across the basin. There are other smaller fractures and faults to the south of the basin that cause displacements in the aquifer, but despite these, research has concluded that the Permo-Triassic sandstone of Shropshire acts as a single aquifer and hence is generally in hydraulic continuity (LOCAR, 2000).

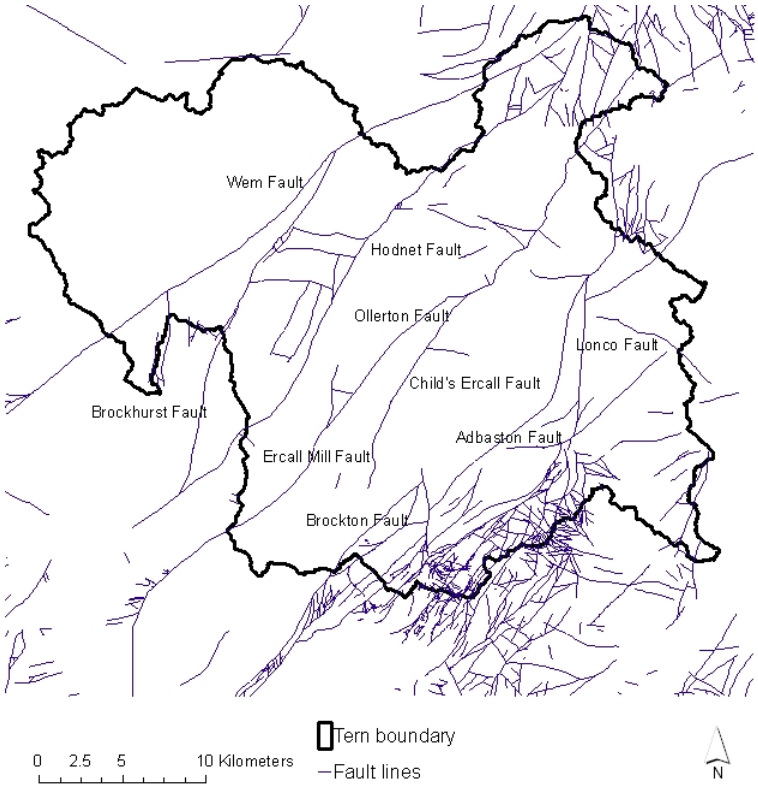


Figure 3.4 Fault-lines in the Tern Catchment

3.3.2. Solid geology and Permo-Triassic stratigraphy

Figure 3.5 highlights the extent of the sandstone outcrop as well as the outcrop of other formations such as the Mercia Mudstone group, Penarth and Wilkesley halite formation which have low permeability in contrast to the highly permeable sandstones. These low permeability formations are mostly contained to the Cheshire Basin and the eastern margins of the Tern catchment in the Stafford basin. The Eturia and Salop and Halesowen formations are interspersed within the Sherwood Sandstones, with these groups are classed as semi-permeable in nature in comparison to the sandstone and mudstone groups (NRFA, 2008).

The Permo-Triassic Sherwood Sandstone and Mercia Mudstone that underlay the Tern catchment can be sub-divided into five formations, each characterised with different hydraulic properties and different textural and visual appearance. Figure 3.5 shows the spatial distribution of the different Sherwood Sandstones in the Tern Catchment and Table 3.2 provides further detail of the description, location and depth of the formations.

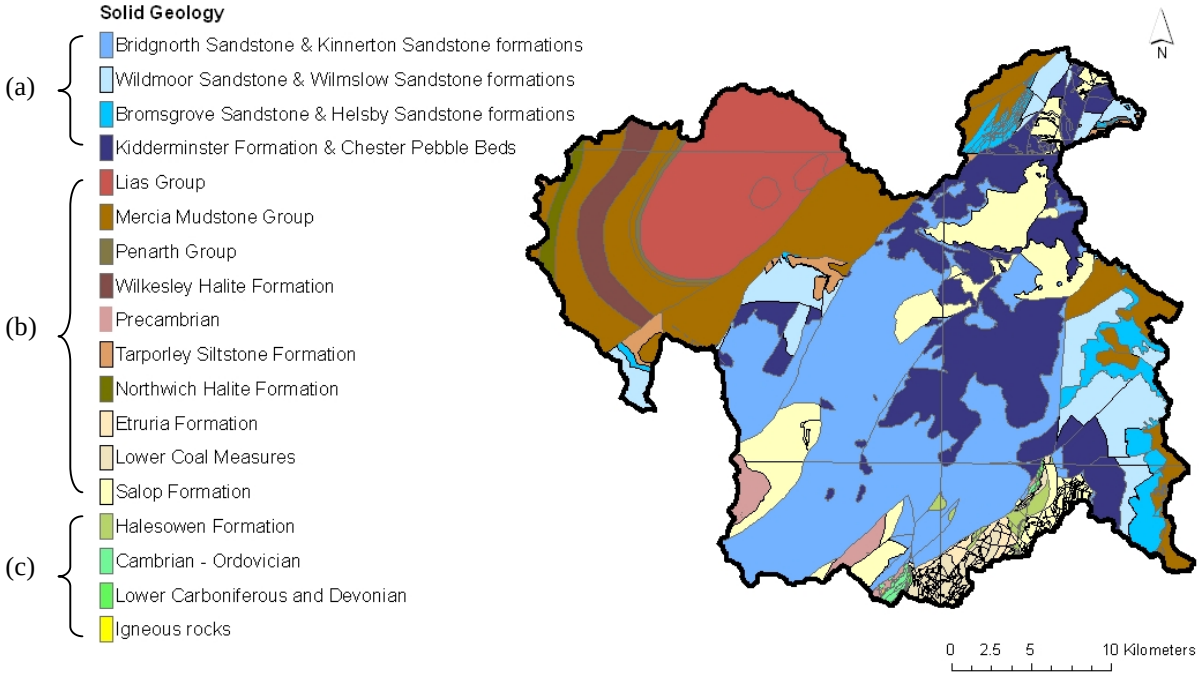


Figure 3.5. Solid Geology of the Tern Catchment (a) formations in the permeable Sherwood Sandstone group, (b) Mudstones & formations of low permeability, (c) formations of mixed permeabilities

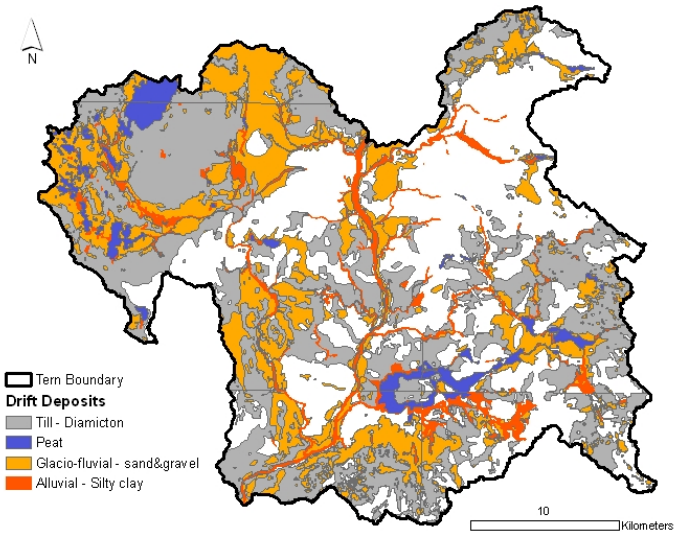
Table 3.2. Stratigraphy of Sherwood Sandstone and Mercia Mudstone. Compiled from (LOCAR, 2000)

Formation name	Description	Location and depth
Bridgnorth sandstone and Kinnerton sandstone formations	Wind blown dune sands of high porosity. Characterised by red to red-brown, fine to medium grained, pebble-free, dune-bedded sandstones.	Over most of the outcrop of this group it is in excess of 100m deep with a uniform distribution. Prevalent between the Wem fault and Stafford Boundary Also present at deeper depths beneath the relatively impermeable mudstones in the Cheshire and Stafford basins.
Kidderminster formation and Chester pebble beds	Red-brown sandstones, pebbly sandstones and rare mudstones. Chester pebble beds have a lower porosity than the Kinnerton sandstone and are more cemented, they have a lower hydraulic conductivity.	Outcrops through much of the central Tern catchment and is interspersed with Bridgnorth and Kinnerton sandstone formation. Depth of the formation varies in the borehole records but thickens towards southeast.
Wildmoor sandstone and Wilmslow sandstone formations	Wildmoor sandstone is a finer grained stone with brown to bright red colouring with inter-bedded silty horizons. Pebbles are uncommon. Characteristically soft and poorly cemented. Similar nature to Kinnerton formation.	Primarily in the east of the catchment, in the Stafford basin, although also found between the Wem and Hodnet fault. Thick superficial drift deposits in this area restrict the delineation of the outcrop. East of Lonco fault it is more prevalent. Depth ranges between 50m at outcrop to 150m east of Gnosall.
Bromsgrove sandstone and Helsby Sandstone formations	Red-brown and yellow and grey with a pervasive calcareous cement. Formation is resistant to weathering and forms prominent scarps across investigation area.	Generally thin – 20m to 50m at outcrop. Located predominantly east of Lonco fault
Mercia mudstone group	Dark reddish-brown mudstones and siltstones. On top of Bromsgrove and Helsby sandstone formation.	200-300m thick on the Ternhill terrace and prevalent in the NW of the catchment in the Cheshire basin.

3.3.3. Drift deposits

The Tern catchment was entirely glaciated in the Quaternary Period and as a result there are many superficial drift deposits on top of the bedrock, especially in the north-west of the catchment. Figure 3.6 a and b map the spatial extent and type of these deposits, as well as the drift thickness, respectively. It can be seen that the majority of the drift is unsorted glacial till which appears patchy in distribution. Glacio-fluvial sands and gravels line the wider floodplains of the Tern and its tributaries, with alluvial silts along the river course. The thickness of the drift deposits vary, with deep deposits in the north-west of the basin, frequently 30-80m in thickness and hence of relative importance to the sub-surface hydrological functioning of the catchment. The central and eastern part of the basin are relatively drift-free, and deposits are often 2-10m in thickness in the south of the catchment.

(a)



(b)

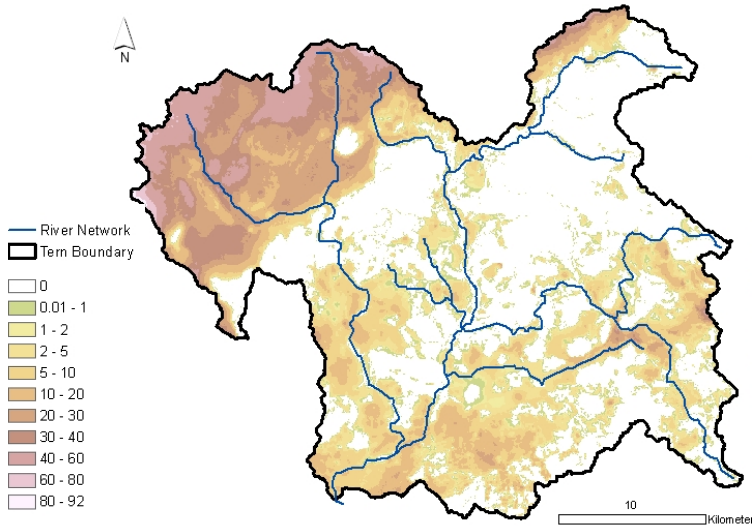


Figure 3.6. (a) Spatial extent and type of drift, (b) drift thickness (m)

3.3.4. Hydrogeology: hydraulic conductivity and anisotropy

The different formations of the Sherwood Sandstone are characterised with differing hydraulic conductivities as summarised in Table 3.3. Research undertaken by Allen (1997) cited in LOCAR (2000) and reviewed as part of the hydro-geological infrastructure of the LOCAR project (LOCAR, 2000), note that data are limited for the Wilmslow formation and there is least data for the Helsby formation.

Table 3.3. Core Hydraulic conductivity of the Sherwood Sandstone Formations (Allen, 1997, summarised by LOCAR, 2000)

Formation	Range ($m\ d^{-1}$)	Inter-quartile Range ($m\ d^{-1}$)	Median ($m\ d^{-1}$)	Geometric mean ($m\ d^{-1}$)
Helsby Sandstone	$1.5x10^{-3}$ – 2.9	0.560 – 4.10	0.86	0.33
Wilmslow Sandstone	$9.3x10^{-3}$ – 6.0	0.056 – 1.13	1.28	0.70
Chester Pebble Beds	$3.1x10^{-4}$ – 15	0.100 – 3.40	1.50	0.57
Kinnerton Sandstone	$2.0x10^{-5}$ - 15	0.020 – 2.60	0.44	0.16

The Kinnerton Formation has the largest range in hydraulic conductivity, likely a result of the lithological diversity in the Formation. The Formation is coarser at lower levels and 'millet seeded' in appearance compared to the fine grained sandstones in the upper part of the Formation. Geometric means of the Formations show the Kinnerton Sandstones also have the lowest hydraulic conductivity (0.16 m d^{-1}) in comparison to the Wilmslow Formation of which hydraulic conductivity is typically 0.7 m d^{-1}. Chester Pebble Beds have a lognormal distribution – likely a result of the cementation of the Formation (Adams et al., 2003). In accounting for the large range in hydraulic conductivity in this formation, hydraulic conductivity is low when the cementation is well developed and higher when there is less cementation.

Table 3.4. Anisotropy of Sherwood Sandstone and Kinnerton Formation

Formation	Orientation	Geometric mean hydraulic conductivity ($m\ d^{-1}$)
Sherwood Sandstone	Horizontal	0.310
	Vertical	0.140
Kinnerton Sandstone	Horizontal	0.220
	Vertical	0.074

Studies of the anisotropy of the Sherwood Sandstone (LOCAR, 2000) have concluded that there is preferential flow of water horizontally rather than vertically, as summarised in Table 3.4. The anisotropy ratio (the ratio between the geometric mean for horizontal to vertical hydraulic conductivity) is 3:1 for the Kinnerton Formation and 2:1 for the Sherwood Sandstone as a whole.

3.3.5. Bulk hydraulic conductivity

Bulk hydraulic conductivities have been calculated by the BGS for 98 sites within the Shropshire Permo-Triassic sandstones (LOCAR, 2000). Bulk hydraulic conductivity (converted from pump test transmissivity data) is preferable at the catchment scale as it avoids the bias of individual core measurements, whether that be from the borehole depth, the depth to the piezometric surface or the thickness of the confining layer (LOCAR, 2000, Freeze and Cherry, 1979).

Table 3.5. Bulk hydraulic conductivity data for Permo-Triassic sandstones of Shropshire (LOCAR, 2000)

Formation	Range m/d	IQR m/d	Median m/d	Geometric mean m/d
All	0.093 - 46	1.8 – 8.4	3.84	3.94
Chester Pebble Beds	0.330 - 16	1.3 – 8.4	3.86	3.28
Kinnerton Sandstone	0.094 - 25	5.9 – 15.0	10.60	6.94

Bulk hydraulic conductivities aid in assessing the significance of any macropore or fracture flow as individual core calculations may be restricted by small scale, yet dominating layers and lenses of low hydraulic conductivity sediments (LOCAR, 2000). The comparative values of bulk hydraulic conductivity to the core data differ considerably (Table 3.3 compared to Table 3.5). A calculated geometric mean of 6.94 m d^{-1} bulk hydraulic conductivity for the Kinnerton Formation is compared to 0.16 m d^{-1} for the same formation from the core data.

3.3.6. Porosity and storage

The porosity is a measure of the void spaces within a material, and is measured as a fraction or percentage.

$$Ø = \frac{V_V}{V_T} \qquad \text{(Equation 3.1)}$$

Where:

V_V is the volume of void-space, and

V_T is the total or bulk volume of material, including the solid and void components.

Ø denotes the porosity.

Table 3.6. Core porosity of Permo-Triassic Sandstone in Shropshire (LOCAR, 2000)

Formation	Range %	IQR %	Median %	Geometric Mean %
Helsby Sandstone	14.7 – 31.2	19.4 – 28.7	25.6	24.2
Wilmslow Sandstone	17.0 – 31.8	24.5 – 29.3	27.4	26.5
Chester Pebble Beds	12.4 – 33.2	19.4 – 27.1	23.5	23.2
Kinnerton Sandstone	5.5 – 34.0	20.0 – 27.6	23.9	23.2

Porosity has been calculated by the BGS from available core samples. As Table 3.6 shows, the ranges within the different Permo-Triassic formations vary considerably. The largest range, seen in the Kinnerton Formation is likely due to the differing grain sizes (siltstones to millet-seed sands) and the degree of cementation (LOCAR, 2000). The smallest range is found within the Wilmslow Formation, and is also calculated as having the highest mean porosity of 26.5%.

The Specific yield (*Sy*) or the drainable porosity is a ratio or percentage less than or equal to the effective porosity. The specific yield indicates the volumetric fraction of the bulk aquifer volume that the aquifer will yield when all the water is allowed to drain out of it by the force of gravity (Freeze and Cherry, 1979):

$$Sy = \frac{V_{wd}}{V_T} \qquad \text{(Equation 3.2)}$$

Where:

V_{wd} is the volume of water drained
V_T is the total rock or material volume

Table 3.7. Typical values of specific yield (Johnson, 1967)

Consolidated deposits	Average Specific Yield (fraction)
Fine-grained sandstone	0.21
Medium-grained sandstone	0.27
Limestone	0.14
Schist	0.26
Siltstone	0.12

Specific yield values have been calculated from pumping test storage coefficients from 42 sites within the Shropshire Permo-Triassic region, with results from these tests suggesting that the specific yield is 0.19 for the region (LOCAR, 2000). These results compare well to the typical values of specific yield calculated by Johnson (1967), as highlighted in Table 3.7, where fine grained sandstones are typically 0.21.

3.4. Soils

This section describes the distribution, description and analysis of soil within the Tern catchment. Data have been acquired for both NATPMAP Soil associations and series, as well as Hydrology Of Soil Types (HOST) data. Description and analysis of the distribution of Tern soils from the NATMAP data are first given, followed by the introduction of the HOST data and analysis within the catchment.

The soil association NATMAP vector data have been obtained from the National Soil Resources Institute (NSRI) at Cranfield University for the area within and buffering the Tern catchment by 2km. The data are a digitised development of the National Soil Map for England and Wales and is the product of two hundred man years of soil survey work in England and Wales (Avery, 1980). The NATMAP vector data are the most detailed of four versions of the National Soil Map and includes information on the soil series (the basic soil types) that are the components of the soil associations on the National Soil Map.

3.4.1. Distribution and analysis of Tern catchment pedology

Figure 3.7 maps the soil associations present within the Tern catchment. Table 3.8 provides further information on the association types including the geology, description of dominant soils and the names of other dominant series within each association. It is important to note that the associations frequently comprise many soil series, more than the two or three dominant soils listed, and individual soils within each association may differ significantly based on many local variances such as geology and topography.

Table 3.8 also provides further analysis of the soil associations within the Tern catchment including the dominant HOST class of each association (described further in Section 3.4.2), as well as the percentage area each association represents within the catchment. There are 32 associations listed within the catchment, with many only covering a small percentage of the area. Table 3.9 lists the main soil associations that cover more than 2% of the catchment.

The general distribution of soils are patchy in nature, however, Table 3.9 shows are two main groups of soil within the catchment; sandy and porus brown earths (Newport1, Wick1, Bridgnorth, Bromsgrove) that in total cover 22.5% of the catchment, and stagnogley soils of Clifton and Salop associations, the most prevalent that cover 49.7% of the area.

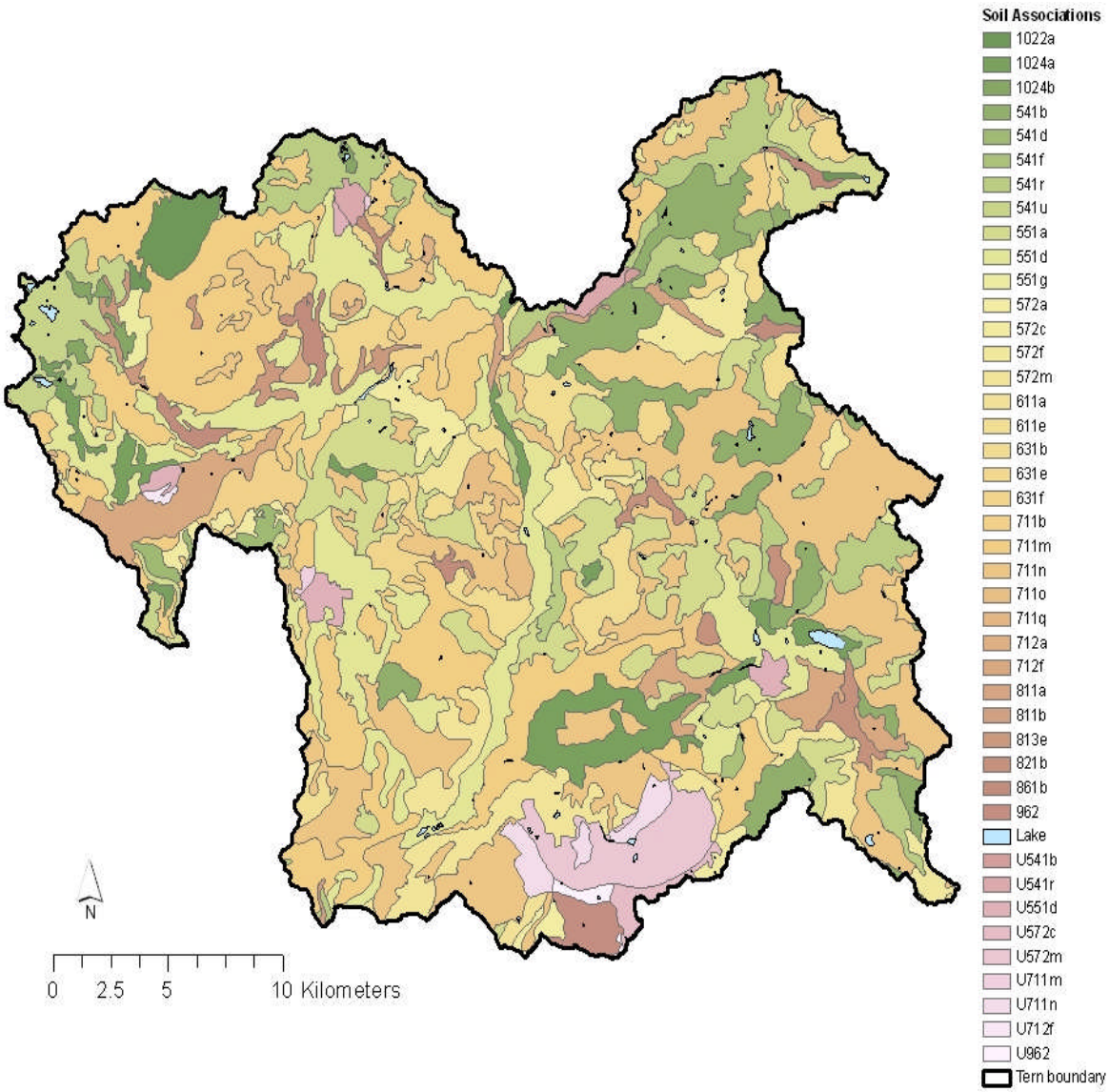


Figure 3.7. Distribution of NATMAP vector soil associations within the Tern Catchment
'Soils Data based on NATMAPvector © NSRI Cranfield University and for the Controller of HMSO, 2010'

Table 3.8. Soil Map Units and associated soil descriptions in the Tern catchment

MAP CODE	MAP UNIT NAME (ASSOCIATION)	SIMPLE DESCRIPTION	GEOLOGY	DOMINANT SOILS	ASSOCIATED SERIES	HOST (dominant)	% Area of catchment
u	URBAN	urban area		Urban area.		-	3.60
lake	LAKE	lake or water body		Lake or water body.		-	0.19
541b	BROMSGROVE	deep loam	Permo-Triassic and Carboniferous sandstone and siltstone	Well drained reddish coarse loamy soils mainly over soft sandstone but deep in places.	EARDISTON HODNET	3	4.29
541d	EARDISTON 2	deep red loam	Devonian and Permo-Triassic sandstone	Well drained often reddish coarse loamy soils over sandstone.	RIVINGTON	4	0.45
541f	RIVINGTON 1	loam over sandstone	Carboniferous and Jurassic sandstone	Well drained coarse loamy soils over sandstone.	MELBOURNE	4	0.06
541r	WICK 1	deep loam	Glaciofluvial or river terrace drift	Deep well drained coarse loamy and sandy soils locally over gravel.	NEWPORT ARROW	5	5.86
541u	ELLERBECK	stony loam over gravel	Glaciofluvial drift	Very stony well drained loamy soils locally on hummocky ground.	BASCHURCH HALL	5	1.71
551a	BRIDGNORTH	sandy over red sandstone	Permo-Triassic & Carboniferous reddish sandstone	Well drained sandy and coarse loamy soils over soft sandstone.	BROMSGROVE CUCKNEY NEWPORT	3	5.70
551d	NEWPORT 1	deep sandy	Glaciofluvial drift	Deep well drained sandy and coarse loamy soils.	WICK RUDGE BLACKWOOD	5	6.03
551g	NEWPORT 4	deep sandy	Glaciofluvial drift	Deep well drained sandy soils.	REDLODGE	5	0.14
711m	SALOP	seasonally wet deep red loam to clay	Reddish till	Slowly permeable seasonally waterlogged reddish fine loamy over clayey, fine loamy and clayey soils.	CREWE CLIFTON FLINT	24	31.20
711n	CLIFTON	seasonally wet deep red loam	Reddish till	Slowly permeable seasonally waterlogged reddish fine and coarse loamy soils and similar soils with slight seasonal waterlogging.	CLAVERLEY QUORNDON SALWICK	24	18.50
711o	RUFFORD	seasonally wet deep red loam to clay	Reddish till and glaciofluvial sand	Slowly permeable seasonally waterlogged reddish coarse loamy over clayey soils.	SOLLOM BLACKWOOD	24	0.35

711q	PINDER	seasonally wet deep loam	Till from Palaeozoic rocks	Slowly permeable seasonally waterlogged fine loamy and fine silty over clayey soils.	PROLLEYMOOR BISHAMPTON	24	1.68
712a	DALE	seasonally wet deep clay over shale	Carboniferous and Jurassic clay and shale	Slowly permeable seasonally waterlogged clayey, fine loamy over clayey and fine silty soils on soft rock often stoneless.	BARDSEY TICKNELL	24	0.12
712f	CREWE	seasonally wet deep red clay	Reddish glaciolacustrine drift and till	Slowly permeable seasonally waterlogged reddish clayey and fine loamy over clayey soils, often stoneless.	SALOP	24	1.55
811a	ENBORNE	seasonally wet deep loam	River alluvium	Deep stoneless fine loamy and clayey soils variably affected by groundwater.	FLADBURY TRENT	10	0.22
811b	CONWAY	seasonally wet deep silty	River alluvium	Deep stoneless fine silty and clayey soils variably affected by groundwater.	FLADBURY CLWYD	9	0.65
813e	COMPTON	seasonally wet deep clay	Reddish river alluvium	Stoneless mostly reddish clayey soils affected by groundwater.	FLADBURY	9	0.37
821b	BLACKWOOD	seasonally wet deep sandy	Glaciofluvial drift	Deep permeable sandy and coarse loamy soils.	OLLERTON FORMBY QUORNDON	10	0.44
572a	YELD	silty over shale	Silurian silty shale. siltstone and limestone	Fine silty soils with slowly permeable subsoils and slight seasonal waterlogging,	BARTON ABERFORD	18	0.67
572c	HODNET	silty over red shale	Permo-Triassic and Carboniferous reddish mudstone. siltstone and sandstone	Reddish fine and coarse loamy soils with slowly permeable subsoils and slight seasonal waterlogging.	STAUNTON MIDDLETON WHIMPLE N.NEWTON	18	1.13
572f	WHIMPLE 3	deep red loam to clay	Drift over Permo-Triassic and Carboniferous reddish mudstone	Reddish fine loamy or fine silty over clayey soils with slowly permeable subsoils and slight seasonal waterlogging.	WORCESTER BROCKHURST	21	2.22
572m	SALWICK	deep red loam	Reddish till and glaciofluvial drift	Deep reddish fine loamy soils with slowly permeable subsoils and slight seasonal waterlogging.	WICK HOPSFORD	18	4.07
611a	MALVERN	stony loam over hard rock	Igneous rock	Well drained very stony loamy soils on moderate to steep bouldery slopes.	MORETON HAMPSTEAD DAVIDSTOW	19	0.15

611e	WITHNELL 2	loam over sandstone	Pre-Cambrian sandstone siltstone and conglomerate	Well drained loamy soils over rock. Sometimes reddish.	BATCH WINSKILL	4	0.23
631b	DELAMERE	sandy over sandstone	Permo-Triassic reddish sandstone	Well drained sandy soils commonly with a bleached subsurface horizon over sandstone.	SHIRRELLHEATH BRIDGNORTH	3	1.64
631e	GOLDSTONE	stony sandy over sandstone	Permo-Triassic and Devonian reddish conglomerate and sandstone	Well drained very acid very stony sandy soils with a bleached subsurface horizon, over conglomerate.	MERCASTON EARDISTON	3	0.46
631f	Crannymoor	deep sandy	Glaciofluvial drift	Well drained sandy soils mostly under woodland and very acid with a bleached subsurface horizon.	NEWPORT BLACKWOOD SOLLOM	5	0.83
711b	BROCKHURST 1	seasonally wet loam to clayey over red shale	Permo-Triassic reddish mudstone and till	Slowly permeable seasonally waterlogged reddish fine loamy over clayey soils.	SALOP WHIMPLE	24	1.64
861b	Isleham 2	seasonally wet deep sand	Glaciofluvial drift and peat	Deep permeable sandy and peaty soils affected by groundwater.	BLACKWOOD OLLERTON	10	0.59
962	NEUTRAL RST OPENCAST	restored following opencast coal working	Carboniferous shale and sandstone and associated drift	Restored opencast coal workings. Slowly permeable seasonally waterlogged compacted fine loamy and clayey disturbed soils. Often stony with thin topsoils.	-	24	0.87
1022a	ALTCAR 1	peat	Fen peat	Deep peat soils with earthy topsoil.	ADVENTURERS	11	0.59
1024a	ADVENTURERS' 1	peat	Fen peat	Deep peat soils.	ALTCAR	11	1.63
1024b	ADVENTURERS' 2	peat	Fen peat over glaciofluvial drift Tertiary and Cretaceous sand	Deep peat soils over variable subsoils, usually sandy sometimes gravelly.	ALTCAR	11	0.17

Table 3.9. Summary of dominant soil associations that cover more than 2% of the Tern Catchment

Association	Symbol	Rank	% area	Soil type
Salop	711m	1	31.2	Surface water gley soil – typical stagnogley soil
Clifton	711n	2	18.5	Surface water gley soil – typical stagnogley soil
Newport1	551d	3	6.0	Brown soil – typical brown sands
Wick1	541r	4	5.9	Brown soil – typical brown earths
Bridgnorth	551a	5	5.7	Brown soil – typical brown sands
Bromsgrove	541b	6	4.3	Brown soil – typical brown earths
Salwick	572m	7	4.1	Brown soil – Stagnogleyic Argillic Brown Earths
Urban soils	Urban	8	3.6	-
Whimple3	572f	9	2.2	Brown soil - Stagnogleyic Argillic Brown Earths

3.4.2. Hydrology of Soil Types (HOST)

The HOST dataset is a hydrologically based classification of UK soils (Boormann et al., 1995). The classification is based on the physical properties and the effects on storage and transmission of soil water displayed by each soil series (of the 1:250 000 England and Wales soil map and corresponding NATMAP data discussed in the previous Section, 3.4.1).

HOST acknowledges the importance that different soils have in governing hydrological response at both local and catchment levels and that the type and distribution of soil within a catchment have a major influence on the hydrological processes taking place within it (Boormann et al., 1995). Consequently, HOST was developed especially for hydrological studies with many applications in improving estimates required for low flow and flood estimation methods. Prior to HOST it was difficult to extract the hydrological response of different soil types, thus HOST has the potential to be a useful tool in parameterisation of the soil components of catchment hydrological models (Lilly et al., 1998; Dunn and Lilly, 2001; Marechal and Holman., 2005).

The classification was derived by assessing hydrological differences between soil classes using regression analysis against long term flow data in 800 catchments in the UK. Following this, 11 conceptual models were derived and further sub classified into 29 response pathways according to flow and storage characteristics. The HOST dataset

comprises of coverage of the UK on a 1km grid with representative HOST classes for each cell. The Base Flow Index (BFI) and Standard Percentage Runoff (SPR) are the two indices calculated to differentiate different HOST classes.

Figure 3.8 maps the dominant HOST classes within the Tern Catchment on a 1km × 1km grid. Of the 29 existing classes, 10 are dominant within the Tern Catchment, as summarised in Table 3.10. Classes 3, 5 and 24 are the most dominant and correspond to the brown earth freely draining soils and slowly permeable stagnogley soils detailed in Section 3.4.1. The standard percentage runoff values for these contrasting dominant classes indicate that the freely draining soils (classes 3 and 5) have a low percentage runoff between 2-12% and the slowly permeable soils (classes 18, 19, 21 and 24) have a much higher percentage runoff of between 40-47%.

Class 18, a class of perched over slowly permeable substrates are primarily found in urban areas such as the southern part of the catchment where Telford is located, shown in Figure 3.8. Classes 9, 10 and 11 are less prevalent in the Tern catchment, but are characteristically of medium hydrological response with regard to standard percentage runoff, between 25-35%.

Table 3.10. Characteristics of HOST classes and percentage areas in the Tern Catchment

HOST class	Description	% area	Rank	SPR %
3	Permeable, free draining soils on permeable sandy, gravelly, chalk or limestone substrates with deep groundwater (below 2m depth).	18.4	3	2-12
4	Permeable, free draining soils on hard but fissured substrates (including karst) with deep groundwater (below 2m depth).	1.0	8	20
5	Permeable, free draining soils on permeable sandy, gravelly, chalk or limestone substrates with deep groundwater (below 2m depth).	21.7	2	2-12
9	All soils with shallow groundwater (within 1m depth) and artificial drainage.	0.4	9	25-35
10		1.5	7	
11		4.0	5	
18	Slowly permeable soils with slight seasonal waterlogging ('perched' water) over slowly permeable substrates.	7.3	4	47
19	Permeable, free draining soils with moderate storage, over hard impermeable substrates at between 0.5 & 1 m depth.	0.1	=10	45
21	Slowly permeable soil with prolonged seasonal waterlogging ('perched' water) over slowly permeable substrates.	2.2	6	40-47
24	Slowly permeable soil with prolonged seasonal waterlogging ('perched' water) over slowly permeable substrates.	42.3	1	40-47
98	Lake.	0.1	=10	

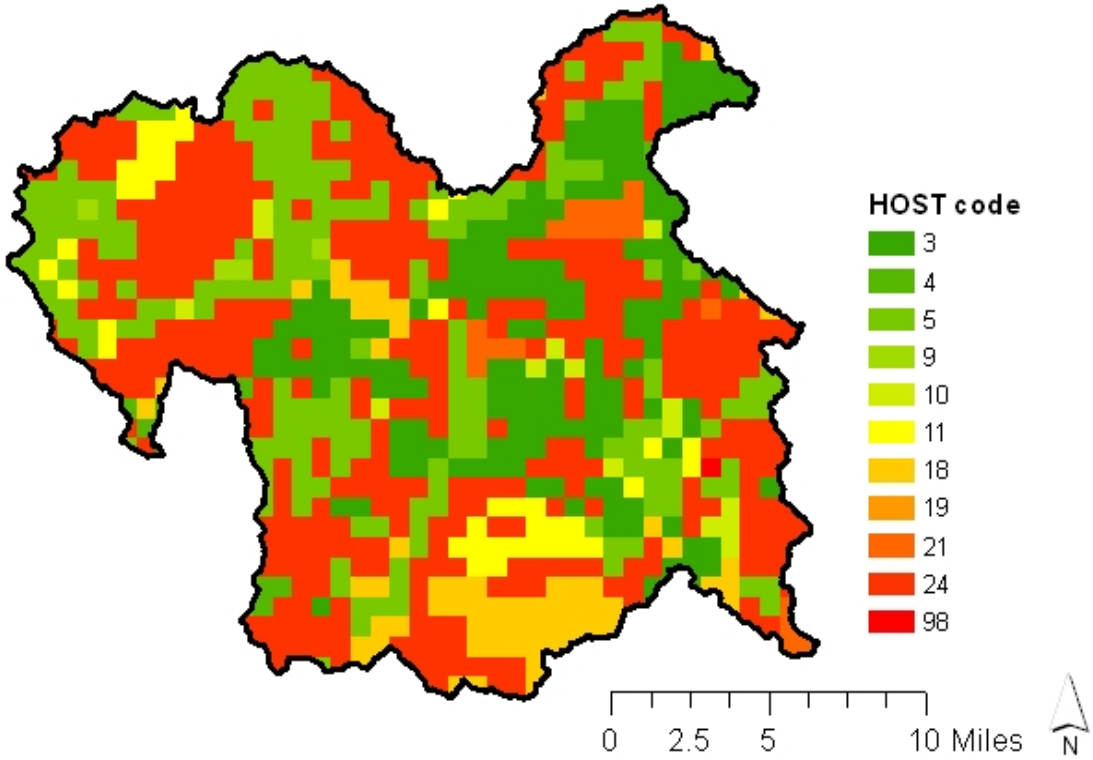


Figure 3.8. Dominant HOST classes on a 1km grid in the Tern Catchment

3.5. Land-use

Land Cover Map (LCM) 2000, 25m raster data were obtained from the Centre of Ecology and Hydrology, Wallingford. The land use data for the Tern catchment are mapped in Figure 3.9. The LCM2000 raster data of the UK are a supervised maximum likelihood thematic classification of spectral data recorded by Landsat TM data and supported by independent ground reference data, to add context and refine the spectral classification (Fuller et al., 2002). 88% of the LCM2000 classification has been derived using multi-temporal summer and winter images to improve the classification accuracy.

The raster data were created by converting the land parcels within the available vector dataset and re-sampling to a 25m grid. Each 25m cell represents the dominant vegetation cover. The LCM2000 data classifies land cover in the UK into 26 sub-classes of which 13 are present within the Tern catchment. As indicated in Table 3.11, the 13 sub-classes have been further refined for this research by merging some of the similar sub-classes together in ARCGIS to give eight predominant classes within the basin. The area and percentage basin cover for the reclassified eight classes are also given in Table 3.11.

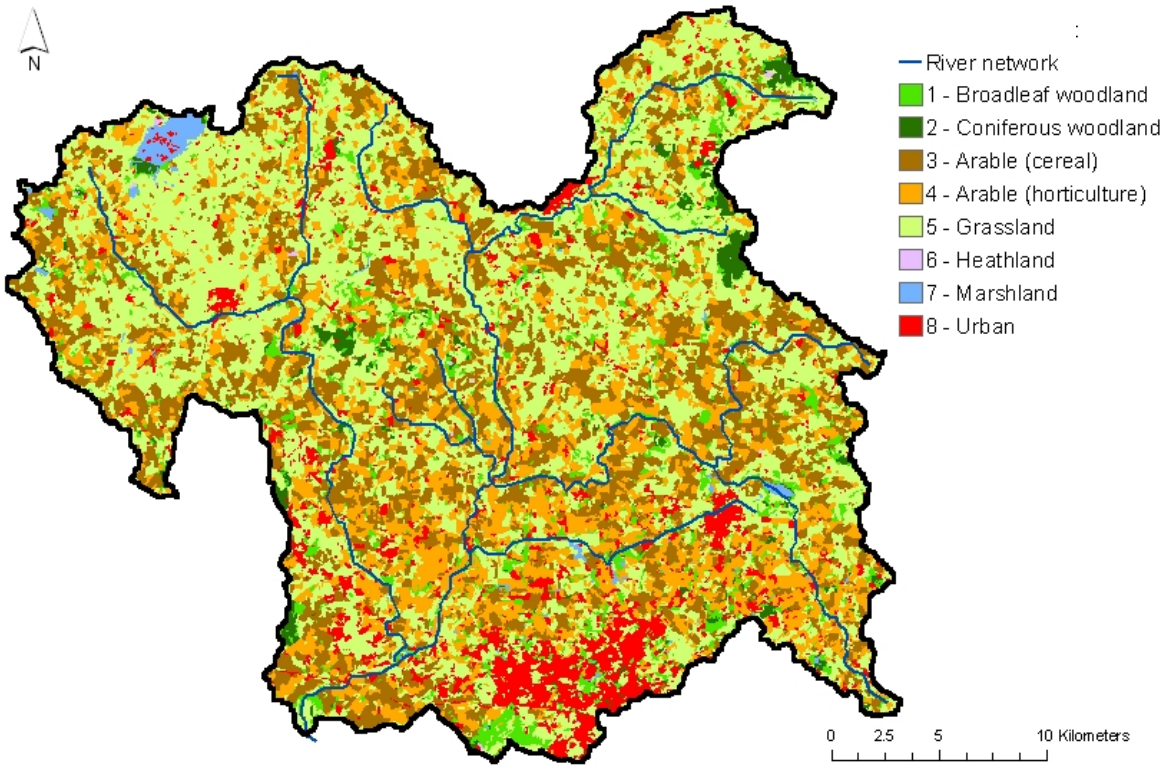


Figure 3.9. Land use within the Tern Catchment

As depicted in Figure 3.9, and shown specifically for each land use class in Figure 3.10, a-h) the predominant cover class is grassland covering 40.8% of the catchment (Table 3.11). Grassland is especially prevalent in the northern half of the catchment, whilst the cereal and horticulture arable classes together contributing 43.4%, are very scattered across the area. 6.9% of the area is designated as broadleaf and coniferous woodland, and is mostly located in the north-east of the catchment in the headwaters of the Tern. The northern section of Telford in the south is the main contributing urban area of 7.6% of the catchment.

Table 3.11 summarises key statistics of the land use distribution, including analyses of land 'parcel' numbers and areas. A parcel is classed as a unit of area designated to one particular type of land use. The smallest unit possible with the data considered is 625m^2 (according to the raster data resolution described above). Parcel analyses confirm the grouped nature of the urban class, with the large individual parcel size of 20km^2.

Table 3.11. Parcel and land use analyses within the Tern Catchment

ID	Name	Area (km^2)	Total area (%)	Total number of parcels	Largest parcel area (km^2)	Mean parcel area (km^2)
1	Broadleaf forest	46.6	5.3	1144	1.6	0.04
2	Coniferous forest	16.5	1.9	249	2.4	0.07
3	Arable (cereal)	200.6	22.9	1140	7.0	0.18
4	Arable (horticulture)	179.4	20.5	1466	6.7	0.12
5	Grassland	357.7	40.8	1106	145.5	0.32
6	Heathland	0.7	0.1	21	0.1	0.03
7	Marshland	8.2	0.9	123	4.2	0.07
8	Urban	66.7	7.6	1005	20.0	0.07

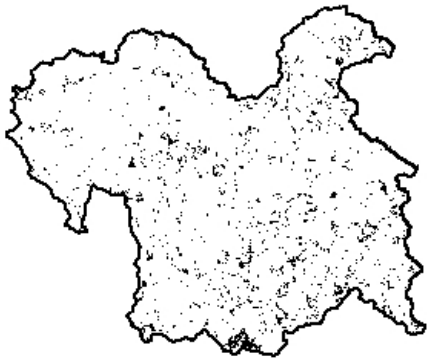
The parcel analysis also indicates that although grassland is the predominant land use in the basin, it does not have the highest number of parcels in the catchment; arable (horticulture), and arable (cereal) as well as broadleaf forest all have more. However, the largest parcel area ($145.5km^2$) is that of grassland, indicating where grassland is located it usually covers a larger unit area, the mean parcel area being 0.32 km^2 – much larger than any of the other cover classes.

Table 3.11, and Figures 3.10 f and g also confirm that marshland and heathland are minor land use classes within the catchment – marshland and small water bodies contributing less than 1% of the total area. However, where present, marshland is clustered into relatively large parcels ($4.2km^2$) – such as Aqualate Mere in the east, and around Ellesmere to the north-west.

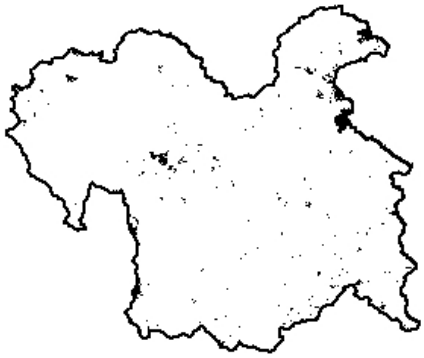
3.6. Catchment meteorology

3.6.1. Precipitation

This section is divided into two parts, the first providing a characterisation of long-term (1975-2000) catchment precipitation in the Tern basin. Analysis of monthly mean precipitation for seven selected gauges is used to highlight seasonal variations within the catchment. Spatial variation in precipitation is also assessed whilst inter-annual analyses are also undertaken to assess any trends in the long-term annual totals which are also compared to the UK average precipitation. Subsequently, the presentation, quality control and analysis of daily precipitation to be used in the catchment modelling (1999-2003) is also provided. The analysis of this recent period is included for the purpose of better understanding the daily precipitation record in the catchment, and ensuring the quality of the data are known before undertaking any modelling.



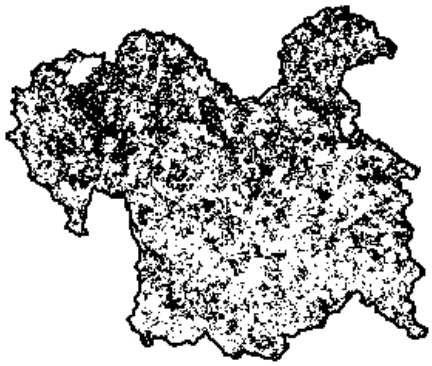
a. Broadleaf forest



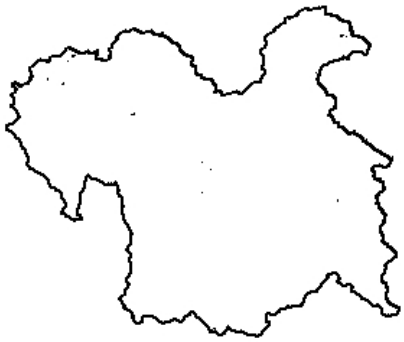
b. Coniferous forest

c. Arable (Cereal)

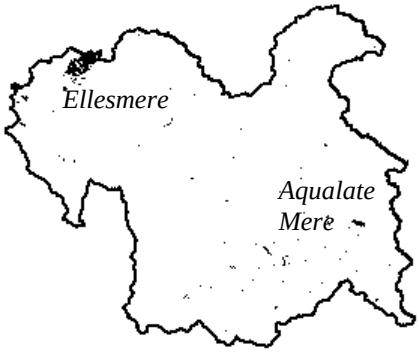
d. Arable (Horticulture)



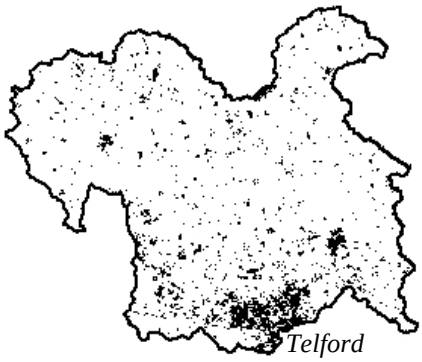
e. Grassland



f. Heathland



g. Marshland



h. Urban

Figure 3.10. Distributions of the predominant land cover classes

3.6.1.1. Characterisation of long-term Tern precipitation (1975-2000)

Substantial precipitation data have been acquired from the Environment Agency and CEH Wallingford for 18 stations falling within or within close proximity to the catchment as illustrated in Figure 3.11. The available time series for each rain gauge are summarised in Table 3.12. It is shown the 1990-2003 period has the highest density of rain-gauges with complete records, with a maximum of 16-17 stations that recorded daily precipitation during this period. Table 3.12 also highlights the specific gauges used within the long-term analysis as well as the gauges used in the next section for the 1999-2003 daily rainfall analysis.

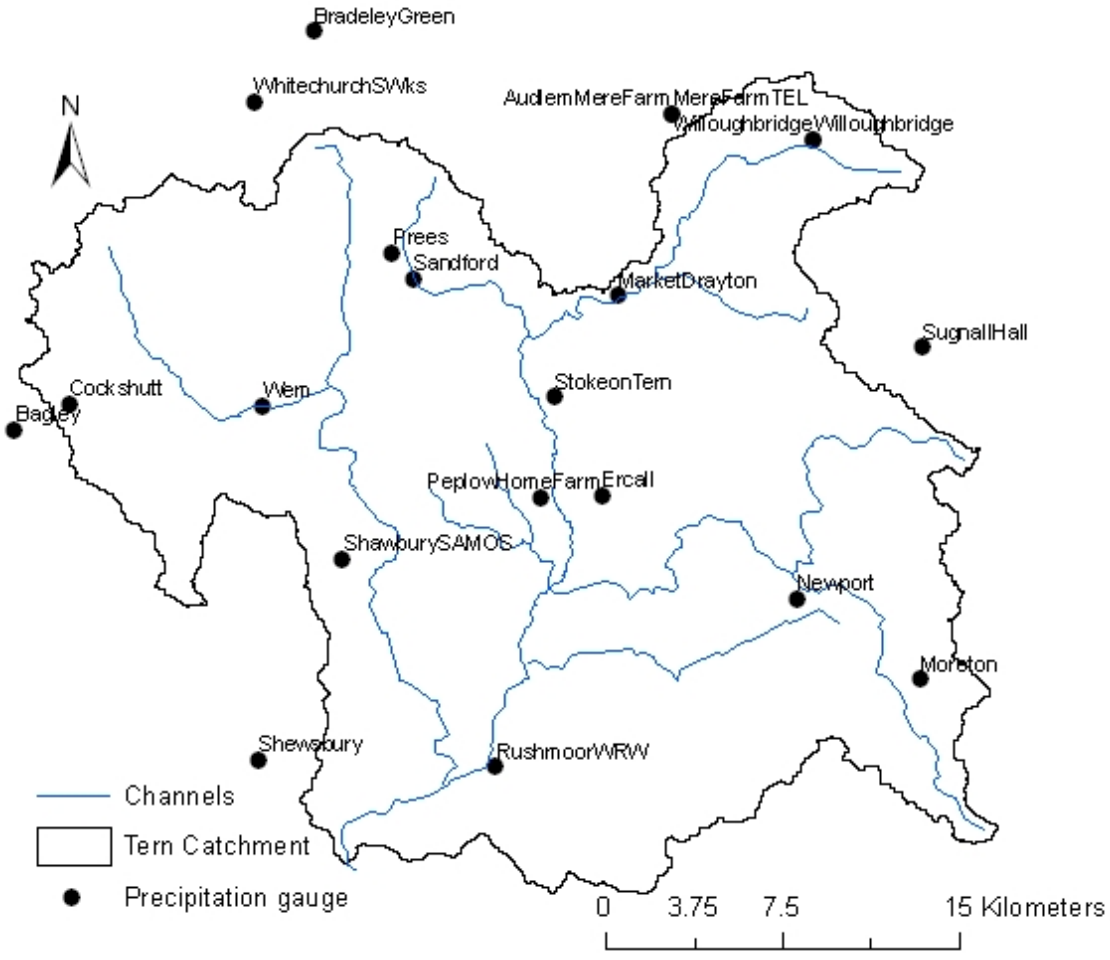


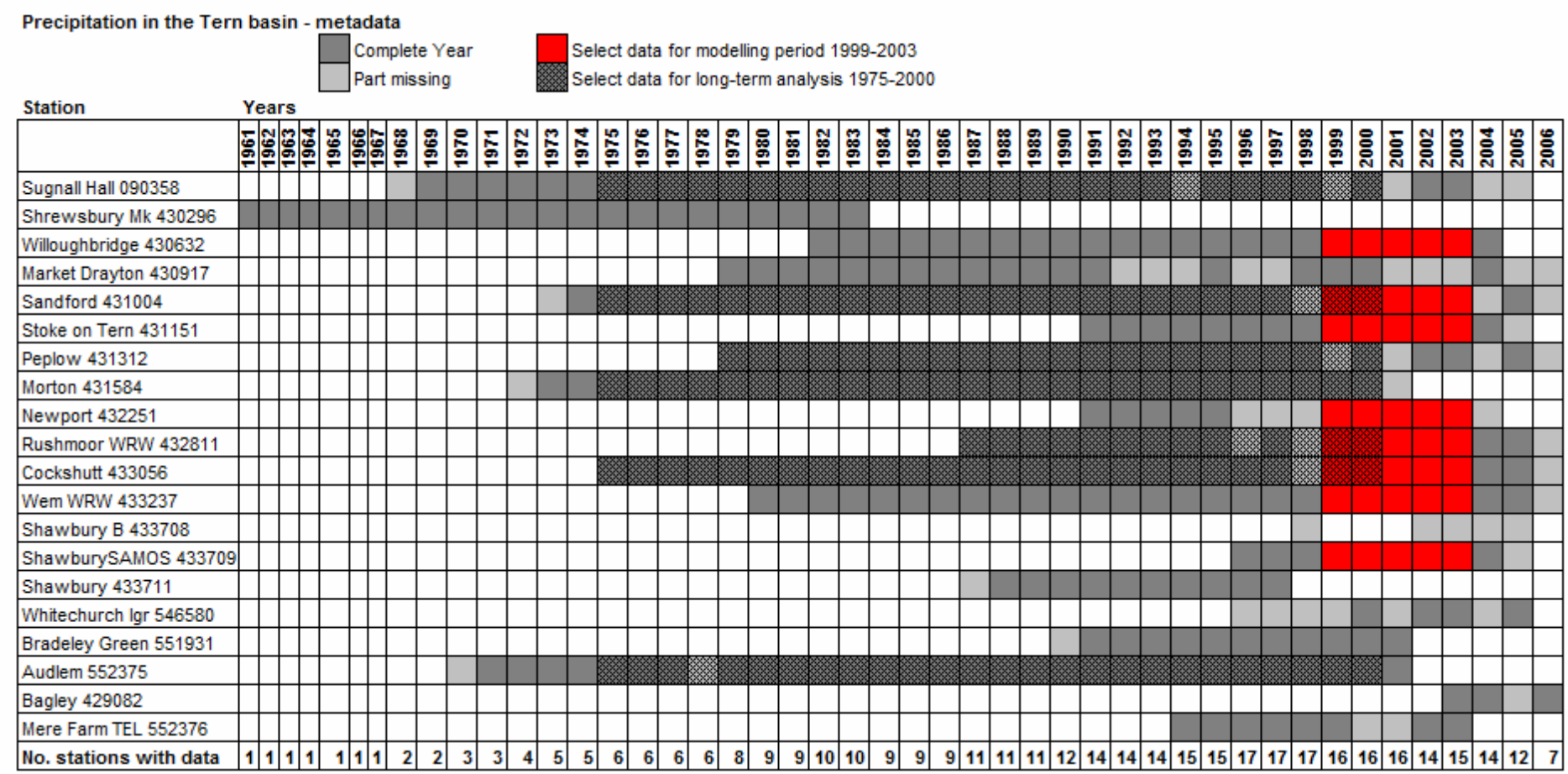
Figure 3.11. Location of rain-gauge sites within and near to the River Tern Catchment

Mean monthly precipitation plots for each station (1975-2000) are presented in Figure 3.12 that characterise the seasonality of precipitation in the catchment. A similar pattern in precipitation between each gauge can be seen, with total monthly precipitation usually not more than 100mm.

December is consistently the wettest month at each station, with a catchment mean of 72.6 mm, and an intra-basin variation between gauges of 13.6 mm. It can also be seen that February (45.3 mm) and July (46mm) are generally the driest months with intra basin variation of 12.7 and 11.1 mm respectively. As the UK has a temperate maritime climate, it is expected that peak rainfall should occur during winter months, and precipitation minima should occur during summer months (Ward and Robinson, 1990), as these data show. The low levels of February rainfall are in-keeping with long-term UK mean monthly trends (Meteorological Office, 2009), likely a result of low levels of moisture in dominant high pressure systems at this time of year.

As shown in the isohyetal map derived in Arc GIS using the spatial analysis Inverse Distance Weighting (IDW) tool (Figure 3.13), the mean annual precipitation ranges from 645 mm yr^{-1} to 750 mm yr^{-1}, with the driest part of the catchment in the south (Rushmoor), and the wettest part of the catchment in the north-east (Sugnall Hall) respectively. The total annual variation across the catchment is thus approximately 100mm yr^{-1}. Table 3.13 and Figure 3.14 relate rainfall and elevation with Figure 3.14 showing a strong relationship between rainfall and the elevation at which the raingauge is situated. The Rushmoor gauge is located at 55.9 mOD (with 643 mm yr^{-1} precipitation), in contrast to the Sugnall Hall gauge at 146 mOD (with 752 mm yr^{-1} precipiation).

Plots of individual gauge total annual rainfall (1975-2000) are shown in Figure 3.15 and confirm the similar pattern in inter-annual variability at all gauges within the catchment. For example, in dry years such as 1975 all of the gauges indicated total annual rainfall of ~500 mm yr^{-1}. Similar total annual precipitation was also seen during wetter years, for example in 2000 where close to 1000 mm yr^{-1} was recorded.



Station	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Sugnall Hall 090358																																															
Shrewsbury Mk 430296																																															
Willoughbridge 430632																																															
Market Drayton 430917																																															
Sandford 431004																																															
Stoke on Tern 431151																																															
Peplow 431312																																															
Morton 431584																																															
Newport 432251																																															
Rushmoor WRW 432811																																															
Cockshutt 433056																																															
Wem WRW 433237																																															
Shawbury B 433708																																															
ShawburySAMOS 433709																																															
Shawbury 433711																																															
Whitechurch Igr 546580																																															
Bradeley Green 551931																																															
Audlem 552375																																															
Bagley 429082																																															
Mere Farm TEL 552376																																															
No. stations with data	1	1	1	1	1	1	1	2	2	3	3	4	5	5	6	6	6	6	8	9	9	10	10	9	9	9	11	11	11	12	14	14	14	15	15	17	17	17	16	16	16	14	15	14	12	7

Table 3.12. Metadata of precipitation records for the River Tern and surrounding area.

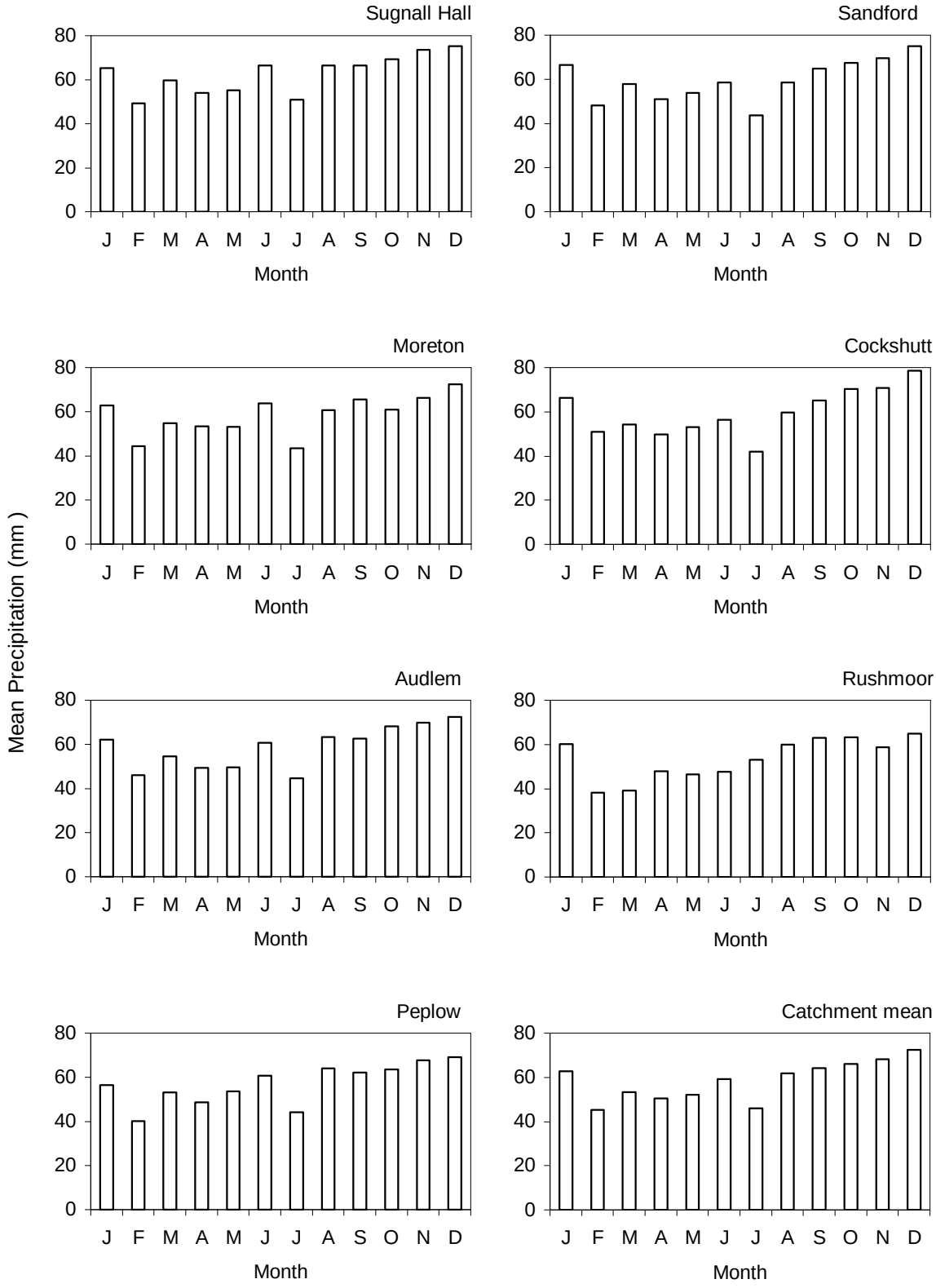


Figure 3.12 Long-term (1975-2000) mean monthly precipitation at individual gauges and the catchment mean

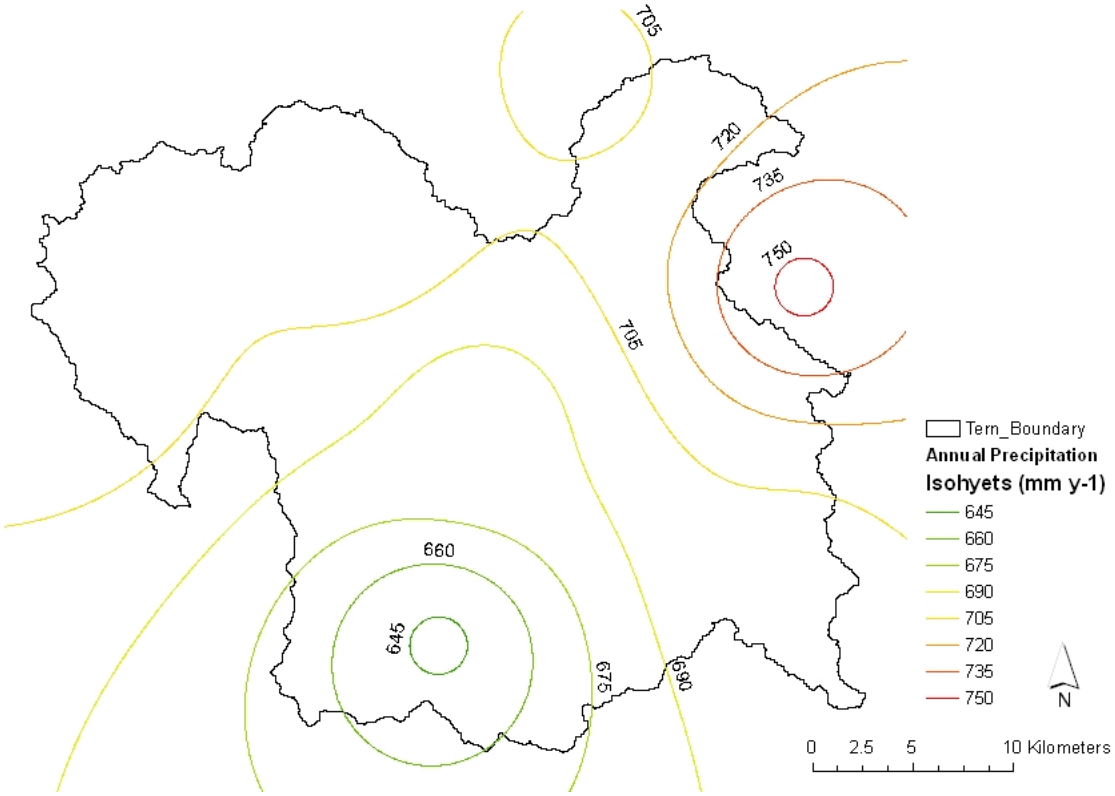


Figure 3.13. Total annual precipitation Isohyet map (1975-2000)

Table 3.13. Mean, maximum and minimum rainfall and elevation of each raingauge

ID	Elevation of gauge (mOD)	Mean annual precipitation (1975-2000)	Max annual precipitation (1975-2000)	Minimum annual precipitation (1975-2000)
Sugnall Hall	146.0	752.0	970.2	504.3
Sandford	86.0	714.8	932.2	515.3
Moreton	90.4	701.5	970.1	493.9
Cockshutt	98.8	717.4	963.0	492.5
Audlem	129.0	703.0	969.3	489.9
Rushmoor	55.9	643.0	839.0	494.9
Peplow	67.8	682.9	883.2	510.9

In context of comparing the catchment rainfall to the UK average, Figure 3.16 shows the long term catchment mean (1975-2003) for the available gauges. Although the 1975-2000 section of the Figure appears to show a general wettening, especially in 1999 and 2000, the year 2003 is shown for both the Tern catchment and UK average that it was a notably dry year, counteracting this trend. As a result, there does not appear to be

any significant drying or wetting pattern within this period, only that total annual precipitation is highly variable between years.

Specific dry years to note include 1975, 1991, 1995 and 1996. In addition, for these years, Tern catchment average rainfall records are notably drier than the UK average. Likewise, for wet years of 1980, 1999 and 2000, the catchment signal appears wetter than the UK average. The catchment rainfall is therefore described as more extreme than the UK average. Despite this, Figure 3.16 shows that the general pattern of inter-annual rainfall is similar in nature.

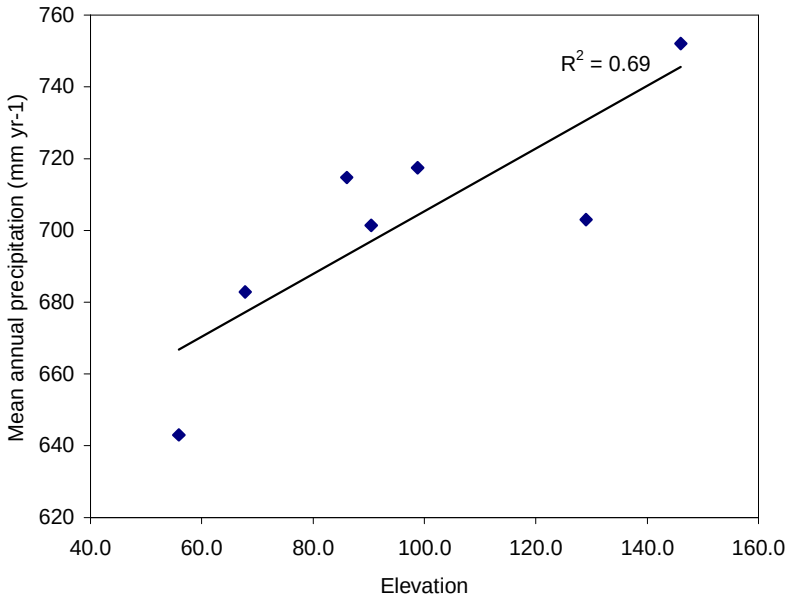


Figure 3.14. Relationship between mean annual rainfall and elevation of gauge

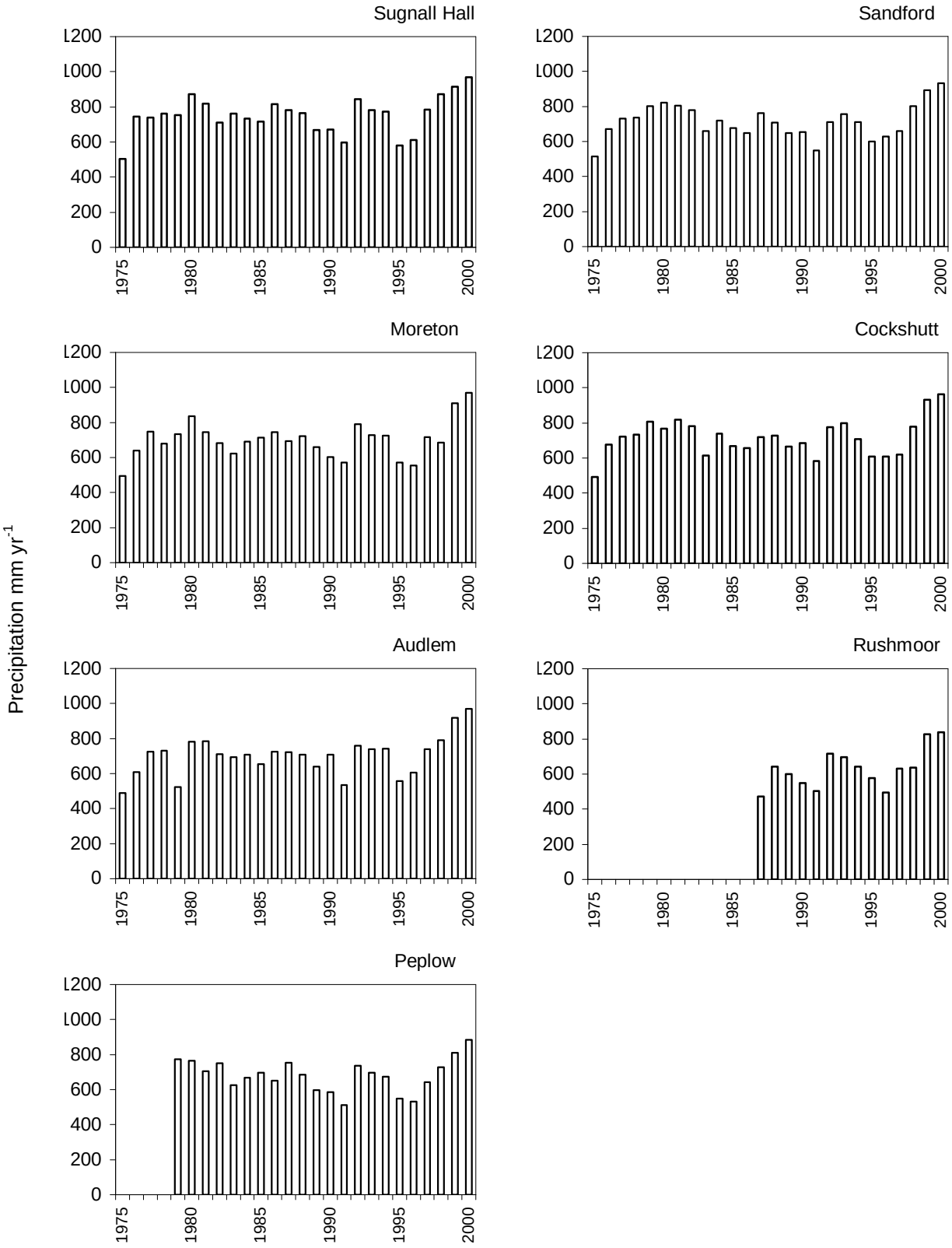


Figure 3.15 Inter-annual rainfall at individual gauges in the Tern Catchment

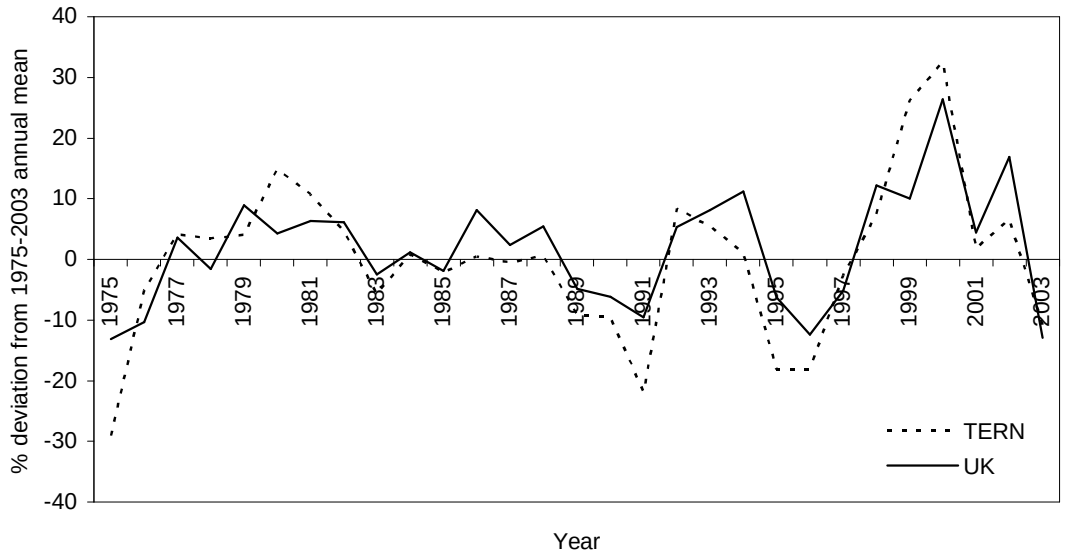


Figure 3.16. Long term precipitation deviation from long term annual Tern mean (1975-2003) and long term average UK precipitation data – percentage deviations from the long term mean. Using data sourced from KNMI, (2008) using the Hadley Centre, Monthly England & Wales precipitation (mm)

3.6.1.2. Precipitation in the 1999-2003 period

As previously introduced, the 1975-2003 rainfall record is not complete for the period in which hydrological modelling is to be undertaken. For this reason, this further daily analysis of rainfall data has been included and, as shown in Figure 3.17 a and b, the spatial distribution of gauges differ as additional gauges are included (Shawbury, Wem, Willoughbridge, Newport, Stoke on Tern) and others not included (Audlem, Sugnall Hall, Moreton and Peplow). However, the distributions are comparatively similar, with both including gauges from across the entire catchment.

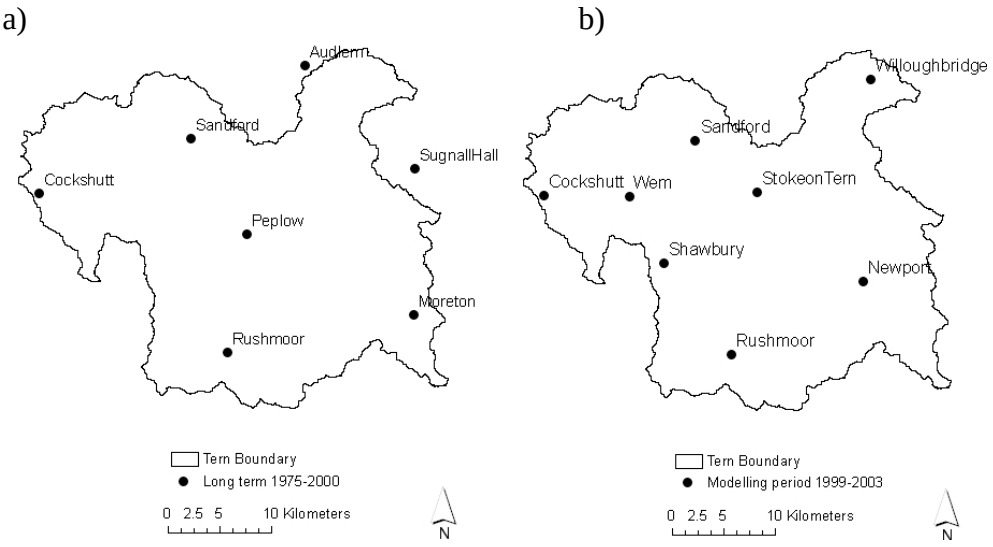


Figure 3.17. a) Spatial distribution of gauges used in 1975-2000 rainfall analyses, b) distribution of gauges for the 1999-2003 analysis and modelling period

Initial quality control of the rainfall data have been undertaken for the gauges with near complete records between the 1997-2003 period. Despite precipitation data having already been quality controlled by the Environment Agency, Double Mass Analyses have been undertaken to further check the data for inconsistencies (Shaw, 1994). The resulting double mass curves (Figure 3.18) plot cumulative gauge data for the questioned gauge against cumulative gauge data from the other nearby gauges. Where deviations in the main gradient of the line were found then the data needed further inspection and analyses for errors and non-continuities (Shaw, 1994). Discrepancies were found for the stations and periods highlighted in Figure 3.19.

The daily rainfall records are complete for all but three of the eight gauges. Figure 3.19 highlights the raw gauged data and missing or suspect periods for Willoughbridge (December 2001), Cockshutt (October 2002) and Newport (November and December 2003). The three stations with periods of missing data (in total four months across three gauges) are still included in the analyses as the missing data periods are minimal in comparison to the additional benefits they bring to the increase in spatial representation of the data in the hydrological modelling. In order to fill the gaps for the daily rainfall records for the three incomplete gauges, different infilling methods of rainfall data were considered. Three different methods of data infilling that are frequently used in hydrological studies include infilling the missing part of the record from the nearest available gauge, or calculating the mean rainfall across an area and using this to infill the missing data (Shaw, 1994). Additionally, gaps in a particular stations record can be in-filled using data from a representative station and by using these data in the profiling equation:

$$\text{Rainfall A} = \text{Rainfall B} * (\text{Average A} / \text{Average B}) \qquad \text{(Equation 3.3)}$$

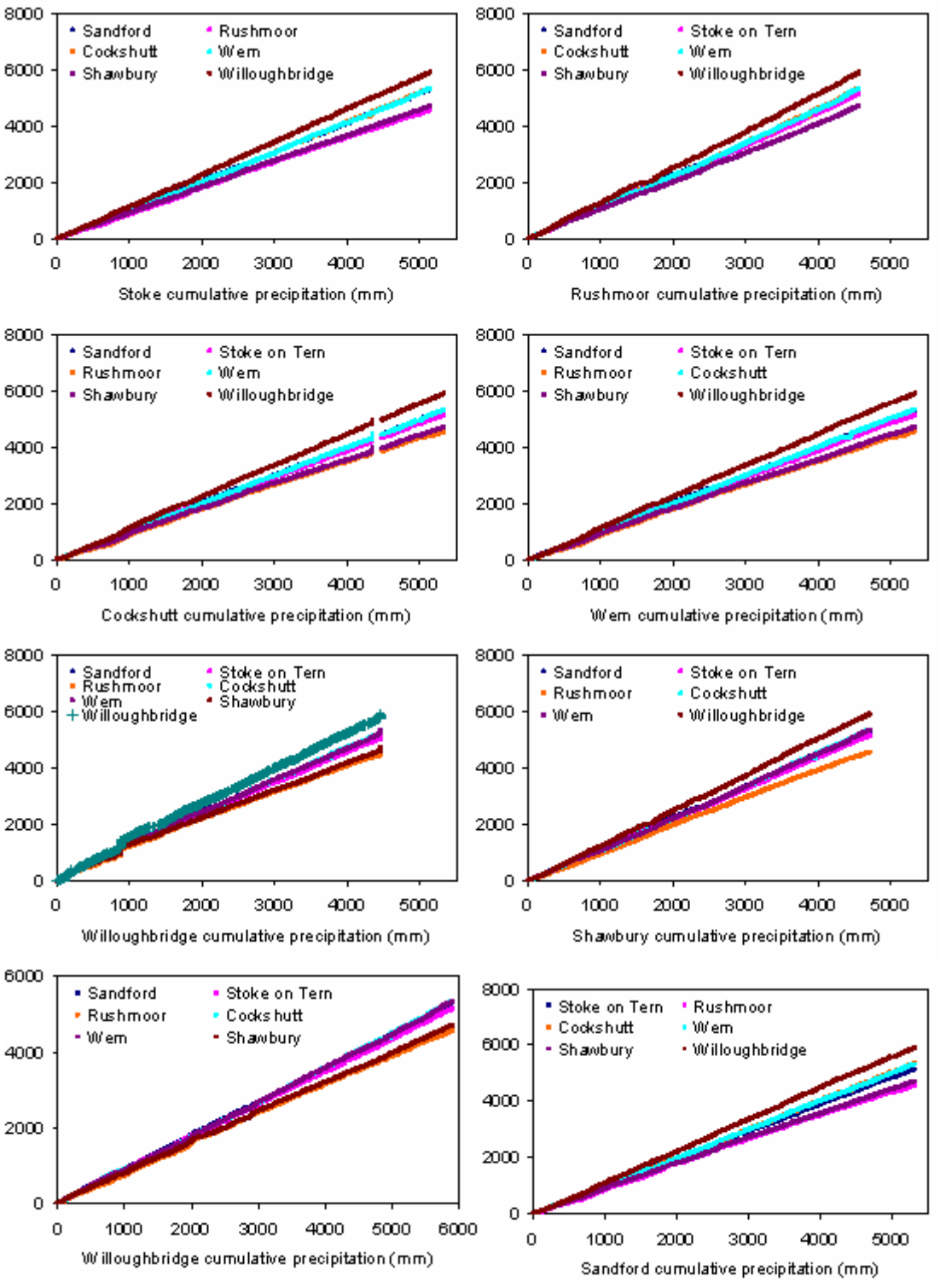


Figure 3.18. Double mass analyses for each gauge for precipitation data quality assessment

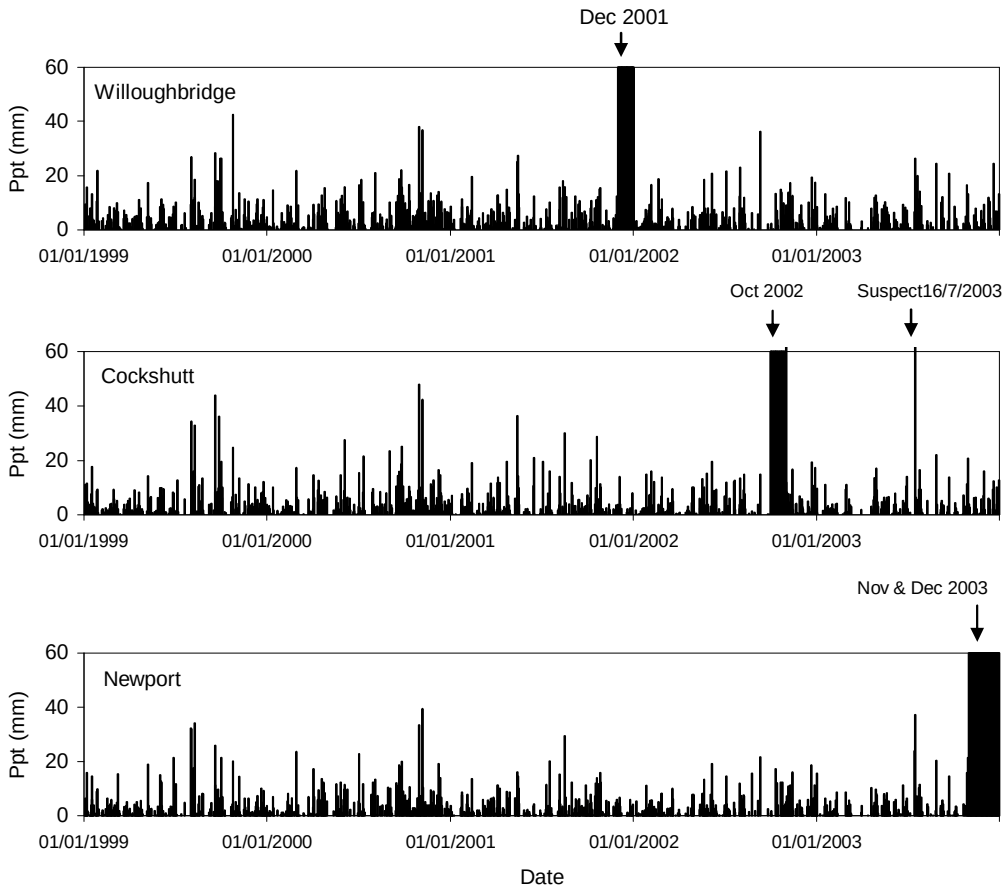


Figure 3.19. Rain gauges within 1999-2003 period with missing data

The mean value used in equation 3.3 depends on the similarity between the two gauging stations; whether they display similar rainfall regimes and whether they are from topographically similar areas. If there is a good similarity between the stations then it is possible to use the long term annual average for the entire record. If the stations show a discrepancy during certain months of the year then it is preferable to use the long term monthly mean. Attention must also be paid to what the in-filled data are to be used for. For detailed hydrological modelling using high resolution data it is important to use detailed average values in the profiling calculation.

As the precipitation data are to be used in a continuous high resolution simulation, the profiling method for infilling the missing rainfall data was selected. As an example of the infilling methodology, the missing November and December 2003 period at the Newport gauge is described:

The rainfall station at Peplow (Figure 3.20) is within relatively close proximity to the Newport station, as indicated in Table 3.14. The two stations are also at approximately the same elevation (difference of 4m). Table 3.14 shows the correlation between the long term records are good at 0.83. Thus, given the similarity of the stations' records, Equation 3.3 has been used to derive new daily data for the missing period.

Table 3.14. Quantitative comparison of stations to be used for profiling and infilling missing precipitation data

Code	Station Name	Missing record	Distance (km)	Elevation (mOD)	Corr R^2	LTA ppt m^{-1}	Profile Ratio	Total ppt m^{-1}
A	Willoughbridge	Dec 2001	6.1	112.5	0.39	85.58	1.12	33.2*
B	Audlem	-		127.0		76.71		29.6
A	Cockshutt	Oct 2002	8.2	96.0	0.82	68.9	0.92	115.6*
B	Wem	-		78.0		73.8		123.8
A	Newport	Nov&Dec 2003	11.8	73.0	0.83	76.7	1.05	115.0*
B	Peplow	-		69.0		73.0		109.3

Where A = Station with missing data, B= Station to be used in profiling
* = calculated value

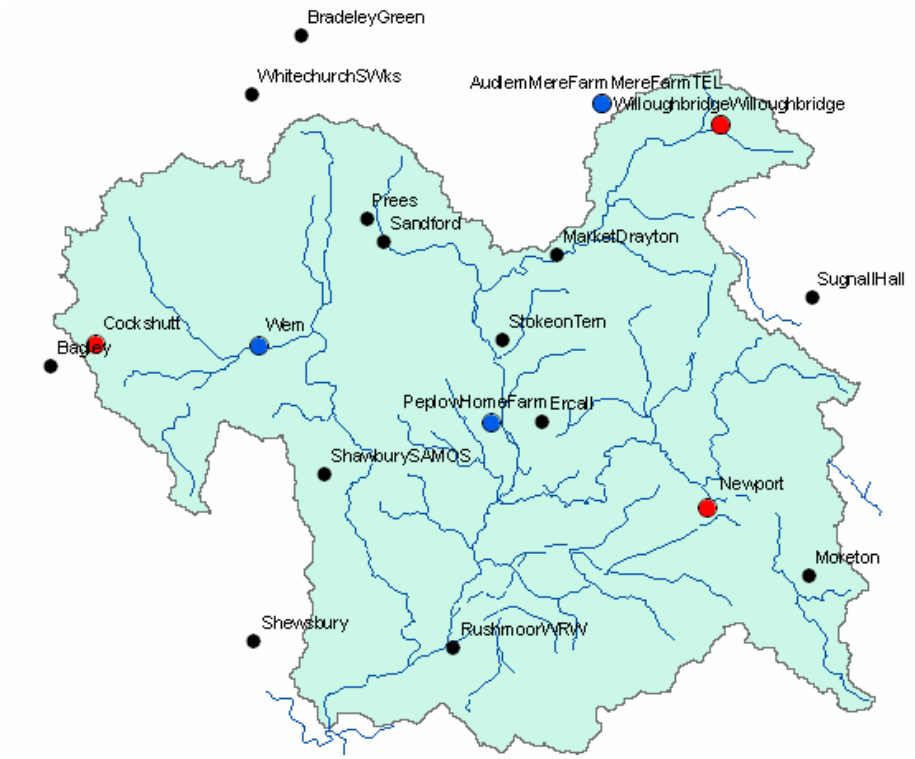


Figure 3.20. Locations of the rain gauges with missing data (red) and rain gauges to be used for infilling (blue)

The same methodology was also adopted for the profiling of Audlem station for the infilling of Willoughbridge (December 2001) data and with the profiling of Wem for infilling of Cockshutt data (October 2002). The distance between the stations, the elevations and statistics comparing the rainfall time series for each station are also given in Table 3.14.

In addition to the Double Mass Analysis and data infilling, Table 3.15 shows the results of a detailed inspection of the daily data. The manual assessment of the data has made it possible to ascertain whether large storm events or significant dry periods were consistent with the other gauges for the same days.

Table 3.15. Suspect daily precipitation records within the modelling period

Gauge Name	**Date**	**Recorded Ppt ($mm\ d^{-1}$)**	**Details (* Recorded Suspect, Environment Agency)**
Cockshutt	06/07/1999	12.8	RS* All other gauges 0 or 0.1mm, could have been a localised event
Newport	02/08/1999	32.2	RS* All other gauges recorded >10mm. This event precedes peak event 03/08/1999
Willoughbridge	07/07/2000	18.4	RS* All other gauges <4mm. Could have been a localised event
Willoughbridge	10/12/2001	6.1	RS* All other gauges <1mm. Could have been a localised event
Newport	25/08/2002	15.6	RS* All other gauges <6.2mm. Event is feasible
Cockshutt	16/07/2003	0.0	All other records between 6.5 – 23.8mm. Co-incides with peak event on the proceeding day (17/07/2003) where 70mm recorded and also suspect.
Cockshutt	17/07/2003	70.0	RS* Far greater than other gauges (41.8mm next closest) and preceding day recorded as 0 (see suspect record above). Likely ppt measurement not taken on 16/07/2003.
Sandford	24/07/2003	0.0	RS* All other gauges 5.3 – 11.2mm. 25/07/2003 Sandford record of 70.3mm indicates likely non-reading of gauge on 24/07/2003
Wem	25/07/2003	49.5	RS* All other gauges 6.7 – 24.2mm (except the suspect Sandford record 70.3mm detailed above).

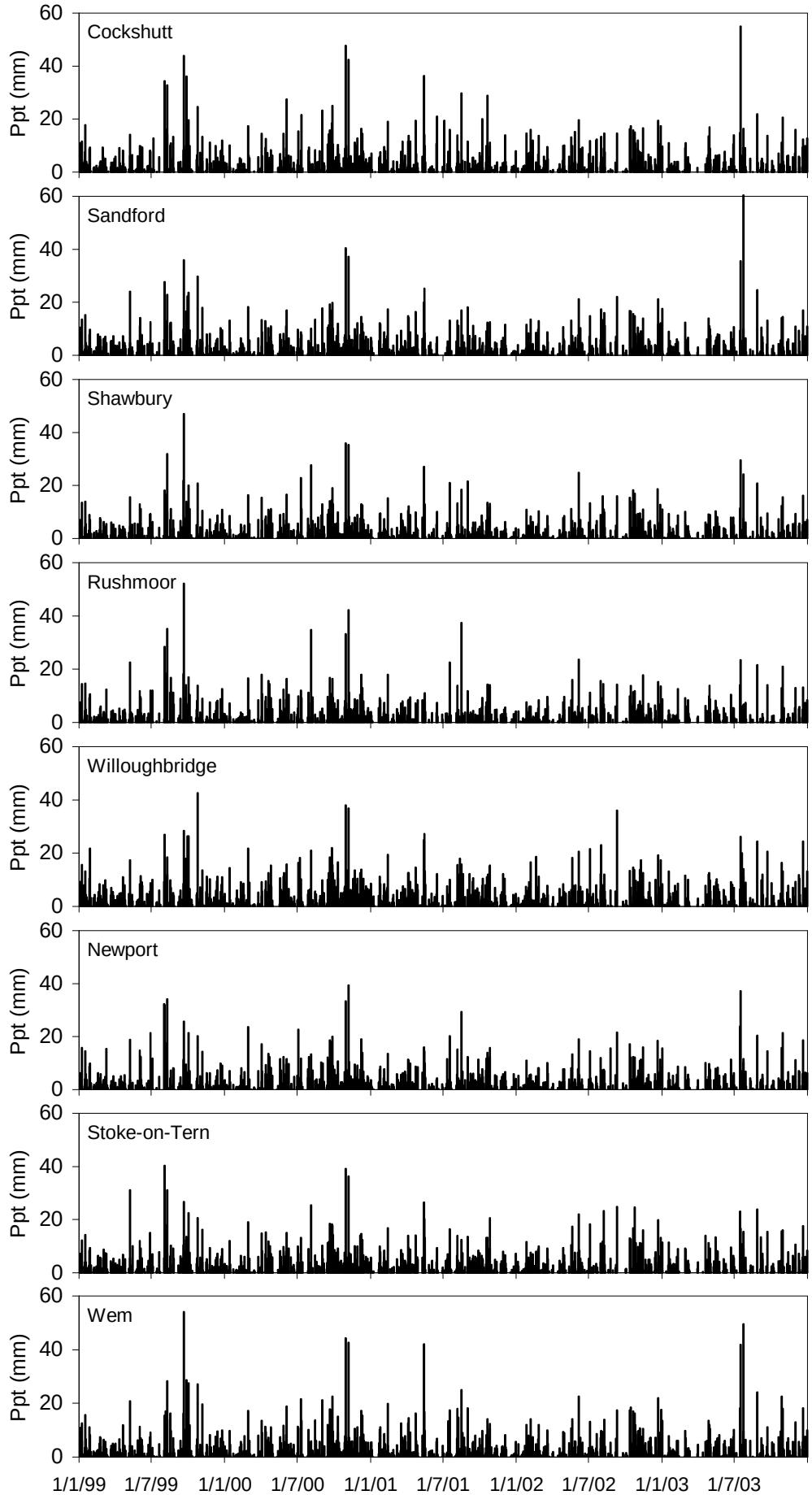


Figure 3.21. Daily precipitation records 1999-2003

As part of the quality control, the suspect rainfall data have been provided in Table 3.15, and the reasons for not altering these data also included. For example, it is possible that although the data appear suspect, they may be a result of individual and localised storm events, such as events 7/7/2000 and 10/12/2001 in Willoughbridge.

The in-filled and inspected daily rainfall data are presented in Figure 3.21 for each gauge station. It can be seen that the main rainfall events occur on the same day and approximately the same magnitude for each station, for example in the winter of 1999-2000. Likewise the dry periods are consistent across the gauges.

In order to re-confirm the already detailed elevation driven rainfall pattern in the catchment, the 1999-2003 gauge distribution has been used to assess if the same pattern is found with the new data. As shown in Figure 3.27, the elevation-precipitation analysis again indicates a strong relationship (R^2=0.86) between elevation of the gauge and recorded precipitation, thus confirming the pattern found in the long term record 1975-2000. As highlighted in Figure 3.27, annual mean precipitation is greatest at higher elevations – with the Willoughbridge gauge in the north-east as already displayed, receiving the most rainfall.

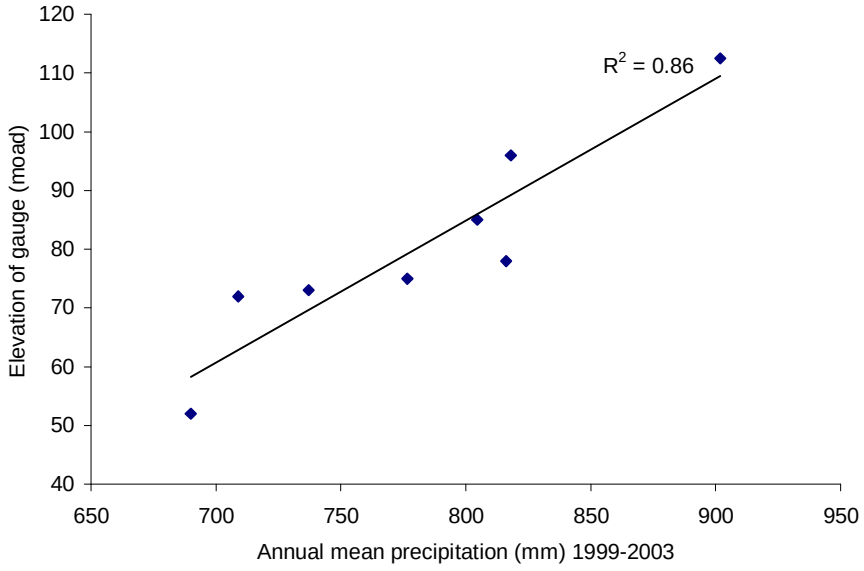


Figure 3.22. Elevation-precipitation relationship for 1999-2003 rain gauge distribution

Table 3.16. Summary statistics of total annual precipitation 1999-2003

Gauge name	Annual mean 1999-2003	Elevation (mOD)	Max (2000)	Min (2003)	Rank (1=wettest)
Willoughbridge	901.9	112.5	1063	723	1
Cockshutt	818.1	96.0	963	673	2
Wem	816.1	78.0	945	697	3
Sandford	804.6	85.0	932	678	4
Stoke-on-Tern	776.5	75.0	923	638	5
Newport	737.0	73.0	896	622	6
Shawbury	708.9	72.0	963	582	7
Rushmoor	690.0	52.0	839	524	8

Details of each of the eight primary gauges are displayed in Table 3.16, with mean annual rainfall 1999-2003, elevation and rank from wettest to driest. The difference in the mean annual precipitation between wettest and driest gauge is 211.9 mm yr^{-1}. Table 3.16 also confirms 2000 as a wetter than average year, and 2003 as a drier than average year.

Section 3.6.1 has reviewed and analysed available precipitation records in the Tern catchment for two time periods. A long term analysis of rainfall data in Section 3.6.1.1 and analysis of 1999-2003 data (the time period for modelling) in Section 3.6.1.2. The long-term analysis has detailed the general precipitation characteristics for the catchment where although intra-catchment rainfall variability was low, a relationship between precipitation and elevation was shown. The purpose of the 1999-2003 analysis and quality control was to better understand the data to be used within the hydrological modelling. The data from both periods appear to show similar trends, with the selected model period highlighted to include both wet (2000) and dry (2003) years for rigorous model testing.

3.6.2. Evapotranspiration

3.6.2.1. Estimating Evapotranspiration

Evapotranspiration is a key component of the hydrological system, with two thirds of precipitation falling on land surface being evapo-transpired globally (Ward and Robinson, 1990). Most of this water is used in plant growth, and hence is important in crop development, ecosystem functioning and consequently many other parts of the hydrological system such as in influencing canopy interception, throughfall to the

surface and throughflow through root networks. Evapotranspiration is the key 'loss' of water from the hydrological system and is more difficult to measure than precipitation and river flow.

Many methods and meteorological formulae have been developed to calculate evapotranspiration (Eto). These range from complex energy balance and mass transfer equations such as the Penman-Monteith (Monteith, 1965) method that requires substantial input data in its calculation including relative humidity and wind speed, to more simple methods requiring minimal data such as the Hargreaves-Samani approach (Hargreaves and Samani, 1982). In addition, field measurement such as by use of evaporation pans were also noted in Section 2.5 as a further method of calculating evaporation.

Oudin et al., (2005) compared the efficiency of 27 different methods of calculating evapotranspiration in four hydrological models. Performance of the hydrological models suggested that it is not always appropriate to use the most complex evapotranspiration equations at the catchment scale. The results of the study therefore provide positive results, especially when calculating Eto in data sparse regions, or where climatological inputs are not available at the necessary spatial or temporal resolution. In addition, the instruments used for the measurement of solar radiation and humidity that are required by methods such as the Penman-Monteith method are often associated with recording errors, such as the observed 1% decrease in measurement of relative humidity accuracy per installed month of operation (Henggeler et al.,1996).

Other more simplistic methods that were included within the assessment by Oudin et al., (2005) required less input data and resulted in equal or better hydrological model performance. The Priestly-Taylor method (Priestly and Taylor, 1972), Thornthwaite (Thornthwaite, 1948), Hamon (Hamon, 1961) and Hargreaves-Samani methods, all demonstrated similar ability to the Penman-Monteith method. Notably, the Hargreaves-Samani (Hargreaves and Samani, 1982), and Thorthwaite methods are driven only with temperature data and solar radiation which make them widely applicable. In this respect, the Food and Agriculture Organisation (FAO) recommends that in the absence of Penman-Monteith, the preferred method to calculate Eto is using the Hargreaves-Samani approach (Allen et al., 1998; Kingston et al., 2008).

With research suggesting the Hargreaves-Samani method being suitable, the following section presents calculation and comparison of the Hargreaves-Samani method for the Tern catchment in Section 3.6.2.1. Additional data are also available from UK-wide modelling of Eto (MOSES-PDM and MORECS). However, as discussed below, the MOSES-PDM data are only available until end of 2001 as the data were derived as a result on an inter-comparison project whereas the modelling period in this thesis extends to the end of 2003, and the MORECS data, discussed below, are mean values of Eto for two 40km^2 grid cells covering the catchment, averaged between 1961-2006. For this reason the MOSES and MORECS data (discussed further in Section 3.6.2.2) are used as comparisons with the calculated Hargreaves-Samani data.

3.6.2.2. Calculation of Hargreaves-Samani evapotranspiration

The most important input parameters required in evapotranspiration calculation are temperature and incoming solar radiation (Oudin et al., 2005). It is noted by Jensen (1985) that 80% of evapotranspiration can be explained by these two parameters. The Hargreaves-Samani approach is an evapotranspiration equation that requires only maximum and minimum recorded temperature and incoming solar radiation by latitude:

$$ETo = 0.0135\ (KT)\ (Ra)\ (TD)\ 0.5\ (TC + 17.8) \qquad \text{(Equation 3.4)}$$

Where:
TD = Temperature max – Temperature min (^{o}C)
TC = Mean daily temperature
Ra = incoming solar radiation (mm d^{-1})
KT = variable coefficient (recommended 0.162 for interior regions and 0.19 for coastal regions, Hargreaves and Samani, 1982)

To calculate Hargreaves-Samani evapotranspiration in the Tern basin, minimum and maximum daily temperature data were acquired from the UK Meteorological Office, MIDAS Land Surface Stations data (1853-current), from the British Atmospheric Data Centre (BADC). DLY 3208 data or the 'monthly return of daily observations' include the required data, and were acquired from three meteorological stations located at three sites within or within close proximity to the Tern catchment (Figure 3.23). Further details of the meteorological stations are given in Table 3.17. The daily time series of minimum and maximum temperature data are given in Figure 3.24.

Figure 3.23. Locations of meteorological stations with daily minimum and maximum data

Table 3.17. Locations and data availability for meteorological stations providing daily minimum and maximum temperature

ID	Name	Location	Available data
651	Newport		01/01/1903 – Present
622	Keele		01/01/1953 – Present
1129	Bradeley Green		01/01/1990 – Present

As Figure 3.24 shows, the pattern of daily minimum and maximum data are similar for all three stations, with annual peaks in temperature during summer (maximum daily temperatures reaching 30^{o}C and minimum daily temperatures 15^{o}C). Temperature minimums occur during winter months, with extreme lows during most years below 0 ^{o}C and as low as -10^{o}C on a number of days. The figures indicate consistent seasonal trends.

Figure 3.24 also indicates missing periods within the data (shown by gaps in the time series). Throughout the 1991 – 2003 period for which evapotranspiration was calculated, the total missing data range from a total of four months at Keele (the most complete record), to 14 months at Bradely Green where there were nine periods of

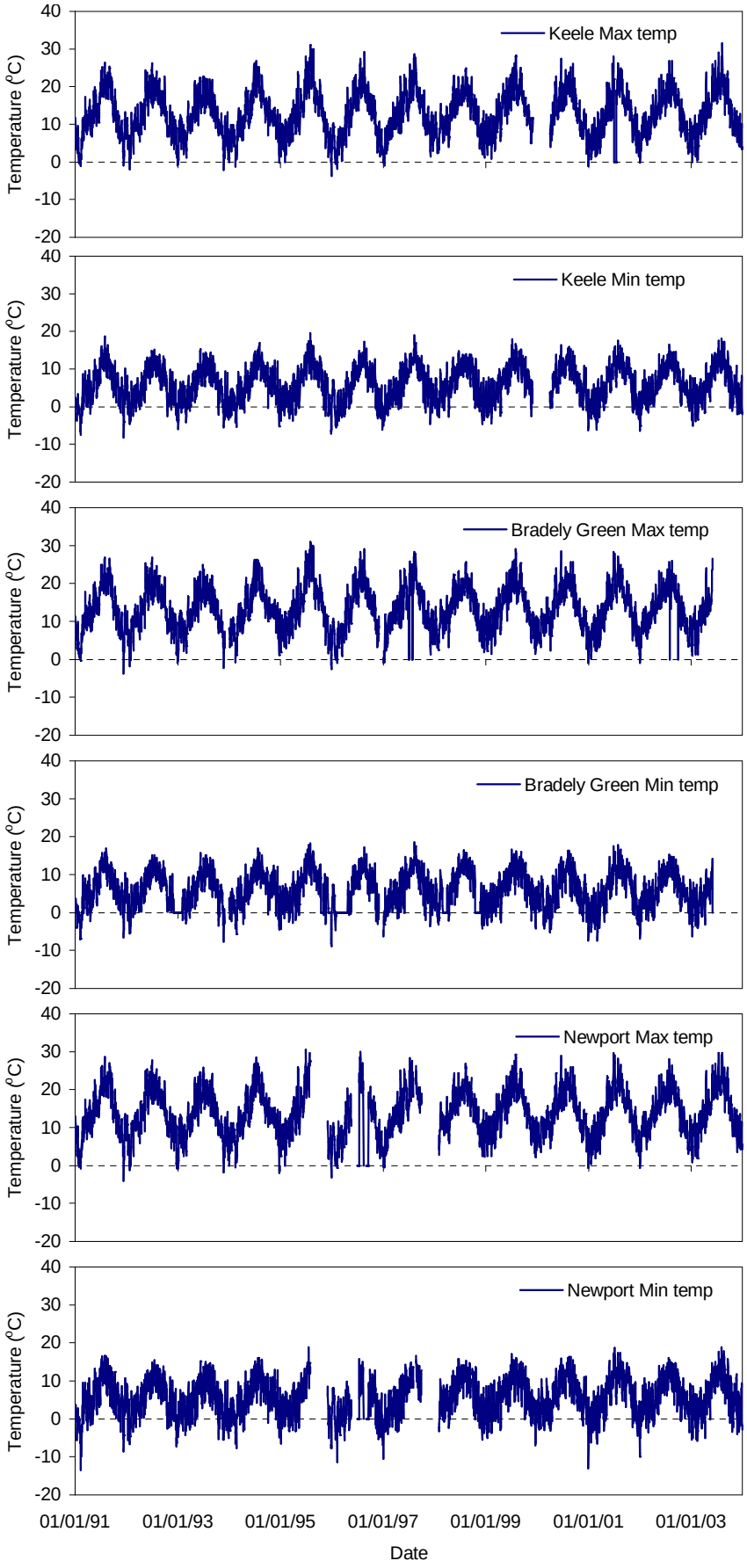


Figure 3.24. Daily time series of maximum and minimum temperature recorded from Keele, Bradely Green and Newport (1991-2003)

missing minimum temperature data. In order to calculate a continuous record of evapotranspiration from these data, interpolation has been undertaken using temperature data from the other nearby stations. Correlation coefficients have been calculated for the relationships between the different sites to determine validity of using nearby records for infilling the data. The R^2 relationships are shown in Figure 3.25, for both minimum and maximum recorded temperatures between each of the sites.

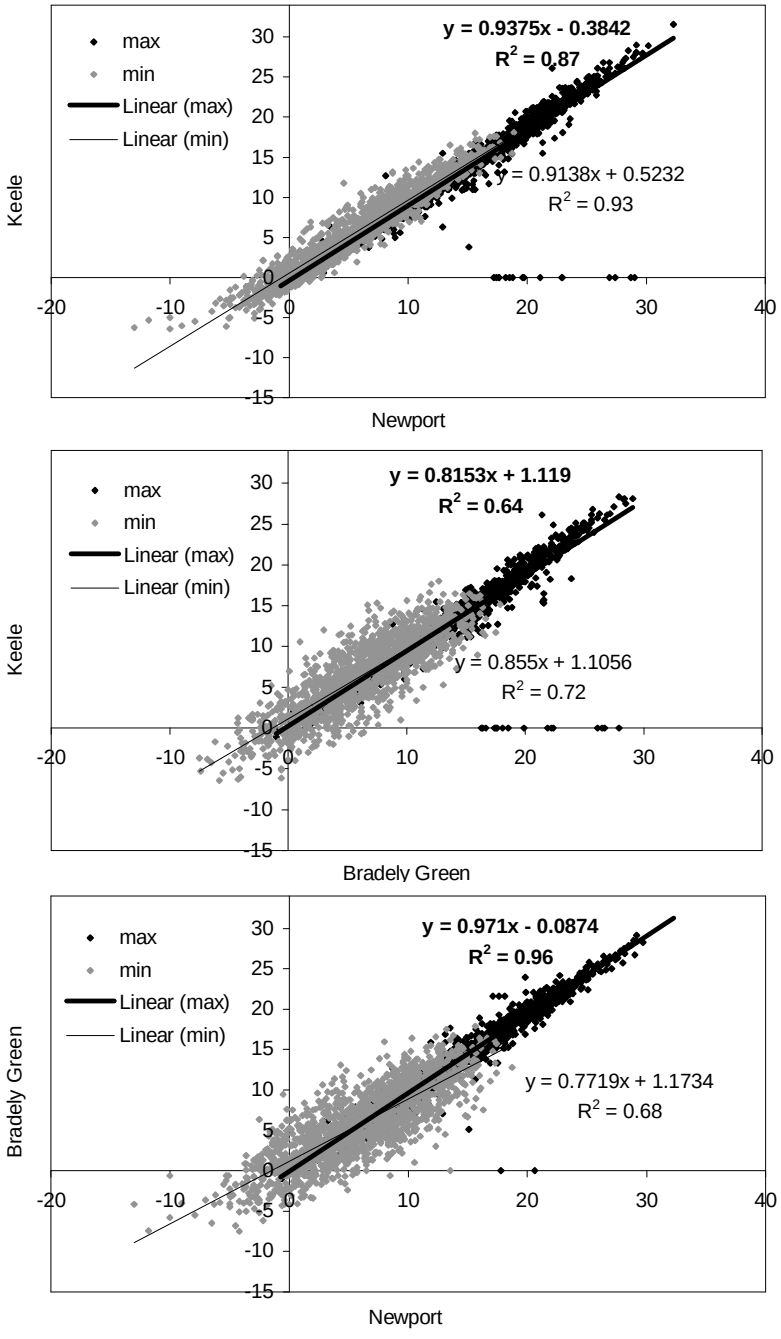


Figure 3.25. Relationships between minimum and maximum temperature at different meteorological stations (axes are all ^{o}C)

Bradely Green and Newport records are the best correlated of all three sites, as signified by the R^2 of 0.96 for the maximum recorded temperatures. Minimum recorded temperatures are less well correlated (R^2 = 0.68) but still show a good relationship, as do the relationships between the other sites, notably for minimum temperatures between Newport and Keele (R^2 = 0.93). Given the good correlation coefficients it is deemed suitable to use a simple infilling method for periods of missing data. Based on the correlation, the missing data were in-filled with the stations indicated in Table 3.18.

Table 3.18. Missing periods of temperature data from the meteorological stations

Missing Period	Maximum temperature	Minimum temperature	Infilling station
1. Newport	Aug-Nov 1995 *02/08/1995 – 01/12/1995*	Aug-Nov 1995 *01/08/1995 – 30/11/1995*	
2. Newport	Mid May-Mid July 1996 *20/05/1996 – 12/7/1996*	Mid May-Mid July 1996 *20/05/1996 – 12/7/1996*	
3. Newport	Mid Aug-Mid Sept 1996 *13/08/1996 – 17/09/1996*	Mid Aug-Mid Sept 1996 *13/08/1996 – 17/09/1996*	Keele, as record is complete for these periods
4. Newport	Mid Aug 1997 *06/08/1997 – 20/08/1997*	Mid Aug 1997 *06/08/1997 – 20/08/1997*	
5. Newport	Oct 1997 – Jan 1998 *02/10/1997 – 01/02/1998*	Oct 1997 – Jan 1998 *01/10/1997 – 31/01/1998*	
Total missing data at Newport	**10.5 months**	**10.5 months**	
1. Bradely Green		December 1992 *01/12/1992 – 31/1/1993*	
2. Bradely Green	December 1993 *2/12/1993 – 1/1/1994*	December 1993 *1/12/1993 – 31/12/1993*	
3. Bradely Green		Part October 1995 *16/10/1995 – 31/10/1995*	
4. Bradely Green		3 weeks in Dec 1995 *7/12/1995 – 26/12/1995*	Keele, as record is complete for these periods
5. Bradely Green		Feb – April 1996 *01/02/1996 – 22/04/1996*	
6. Bradely Green	December 1996 *02/12/1996 – 01/01/1997*	December 1996 *01/12/1006 – 31/12/1996*	
7. Bradely Green		March 1998 *01/03/1998-31/03/1998*	
8. Bradely Green		Mid Oct – Mid Nov 1998 *18/10/1998 – 23/11/1998*	
9. Bradely Green	June – Dec 2003 *02/06/2003 – 31/12/2003*	June – Dec 2003 *02/06/2003 – 31/12/2003*	
Total missing data at Bradely Green	**8 months**	**14 months**	
1. Keele	Dec 1999 – March 2000 *02/12/1999 – 01/04/2000*	Dec 1999 – March 2000 *01/12/1999 – 31/3/2000*	Newport, as record is complete and good relationship between Keele and Newport temperature data
2. Keele	Part July 2001 *02/07/2001 – 17/07/2001*		
Total missing data at Keele	**4.5 months**	**4 months**	

The incoming solar radiation required for the calculation of Hargreaves-Samani evapotranspiration were acquired from FAO estimates of monthly extra terrestrial radiation at 52°N and 54°N (Table 3.19). The mean of the radiation at these two latitudes was calculated to give values for 53 °N, the latitude of the Tern catchment. Table 3.19 also indicates the conversion of units from MJ m^{-2} d^{-1} to mm d^{-1} as required by Equation 3.4.

Derived time series of potential evapotranspiration for each station and the catchment mean are given in Figure 3.26. In calculating the mean catchment evapotranspiration, the minimum and maximum temperatures have first been averaged before being used in the Hargreaves-Samani calculation. General patterns between each station are similar, with Table 3.20 providing comparative statistics. The mean catchment potential evapotranspiration is 1.75 mm d^{-1}, with Newport having the largest mean evapotranspiration of 1.82 mm d^{-1} and Keele the lowest, 1.66 mm d^{-1}. The variation between the different locations is however small, with only a 0.12 mm d^{-1} difference.

Table 3.19. Extraterrestrial solar radiation calculated for 53°N

Month	R_a MJ m^{-2} d^{-1}			*Ra mm d^{-1}
	54 °N	52 °N	53 °N mean	53 °N
J	6.5	7.7	7.1	2.92
F	12	13.2	12.6	5.19
M	20	21.1	20.55	8.46
A	30	30.8	30.4	12.51
M	37.8	38.2	38	15.64
J	41.5	41.6	41.55	17.10
J	39.8	40.1	39.95	16.44
A	33.2	33.8	33.5	13.79
S	23.7	24.7	24.2	9.96
O	14.5	15.7	15.1	6.21
N	7.8	9	8.4	3.46
D	5.2	6.4	5.8	2.39

Values for R_a on the 15th day of the month provide a good estimate (error < 1 %) of R_a averaged over all days within the month. Only for high latitudes greater than 55° (N or S) during winter months deviations may be more than 1%.(Source: Allen et al., 1998)

* MJ m^{-2} d^{-1} can be converted to mm d^{-1}as; mm d^{-1} = MJ m^{-2} d^{-1} / 2.43

Table 3.20. Comparative Eto statistics (Eto mm d^{-1})

	Bradely Green	Keele	Newport	Catchment mean
Minimum	0.08	0.08	0.09	0.10
Mean	1.74	1.66	1.82	1.75
Maximum	5.67	5.72	6.32	5.90

3.7.3. Modelled evapotranspiration data

The Meteorological Office and Environment Agency have developed models that calculate evaporation on a gridded nationwide basis. Modelled evapotranspiration data are available for two methods. The Meteorological Office Rainfall and Evapotranspiration Calculation System (MORECS) calculates the soil state and evaporation using the Penman-Monteith method, for both a 40 km × 40 km grid nationwide on a weekly operational basis, and at individual meteorological stations for hindsight studies. The data are calculated from measurements of temperature, sunshine, wind and humidity, for a variety of soil and vegetation types. The MORECS data for the Tern catchment (Figure 3.27) have been acquired that consist of average monthly values of potential evapotranspiration for a crop type of grass. The data that were available were long term mean values from model runs between 1961 and 2006 averaged for the two $40km^2$ squares 114 and 124 that cover the basin area.

As shown in Figure 3.27, the MORECS mean monthly evapotranspiration displays a normal distribution throughout the year with peaks in summer months and highest value in July (a mean of 2.99 mm d^{-1}). Peak evapotranspiration is approximately six times greater than in December where minimum evapotranspiration is on average 0.51 mm d^{-1}.

In addition to the MORECS data, daily Meteorological Office Surface Exchange Scheme – Probability Distributed Model (MOSES-PDM) archive data are also available from a historical inter-comparison project undertaken in 2003 (Hough, 2003), between MORECS and MOSES-PDM. The inter-comparison project was initially conducted to research the impact of antecedent catchment conditions on flood forecasts and the methods used to estimate these conditions. However, when the wider benefits of a revised system to replace the MORECS data were realised, the project developed to a comparative study between the two data sets and two hydrological events: the 1975-6 drought and the 2000-2001 floods.

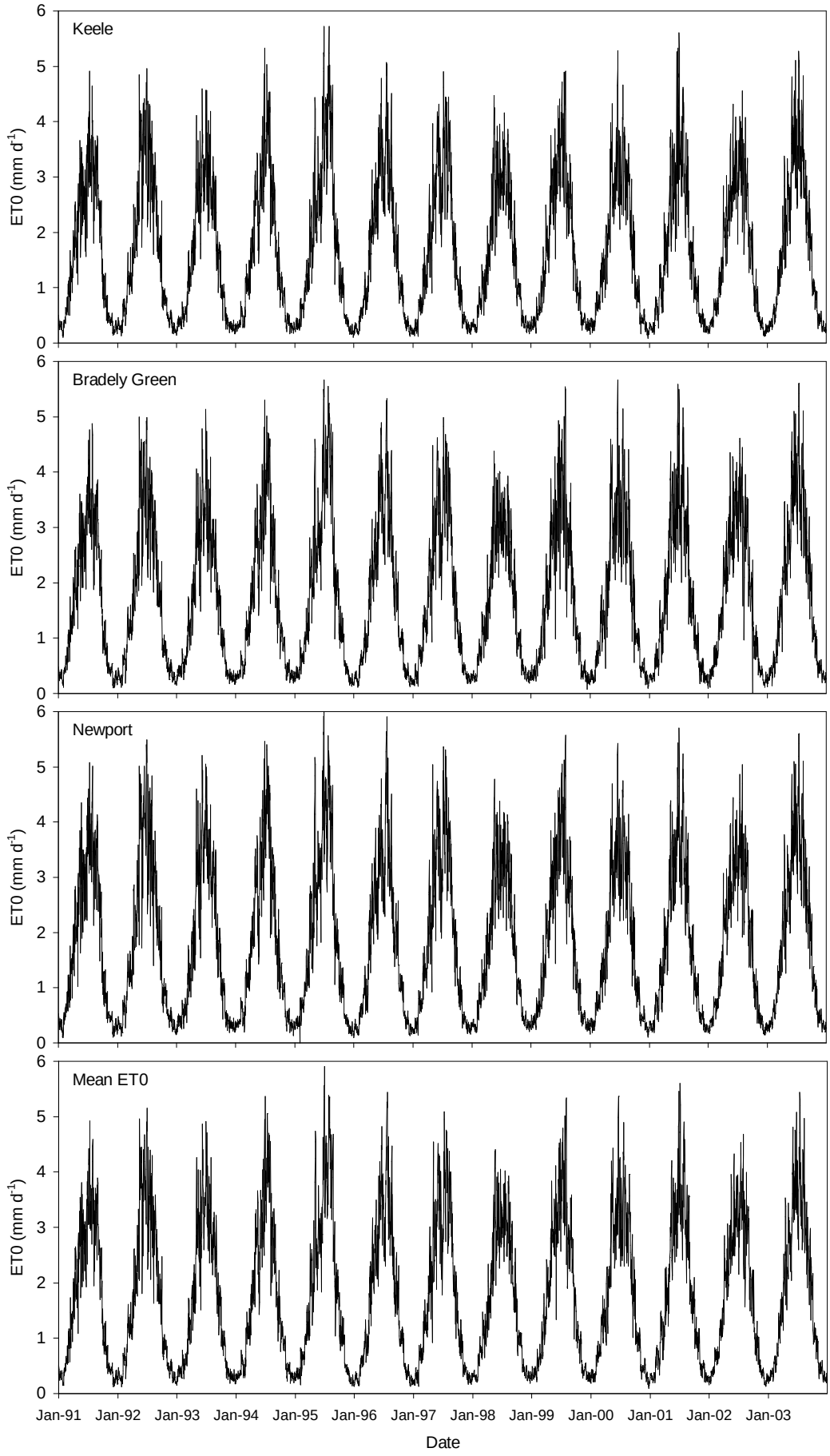


Figure 3.26. Calculated Hargreaves-Samani evapotranspiration for the Tern Catchment

As a result of the project, a daily MOSES-PDM dataset was derived for the UK between 1961-2001. Although operational MOSES data are available at a 5km grid resolution, the data from the historical comparison project (where no suitable historical recorded data were available) use hourly weather data inputs from MORECS 40km × 40km data. Potential evapotranspiration from MOSES-PDM are available for both a uniform grassland scenario and also the appropriate percentage Real Land Use (RLU) covering the 40km × 40km tile. Monthly averaged data for the two tiles covering the basin are shown in Figure 3.27 for the RLU scenario.

It can be seen that there is a reasonable agreement between the averaged MORECS and MOSES data (1999-2001 for when the data are available) although the MORECS data estimates slightly higher PE in all months in comparison to the MOSES-PDM. The Hargreaves-Samani data have also been included in Figure 3.27. It is evident that summer peaks are much larger than for the modelled MOSES-PDM or MORECS data with peak rates in June-August close to 4 mm d^{-1}. The mean daily evapotranspiration rate calculated for the three methods also indicates that the Hargreaves-Samani method has with the highest rate of 1.75 mm d^{-1} compared to a similar 1.62 mmd^{-1} (a difference of 8%) for MORECS (although this is the average monthly 1961-2006 record) and lower 1.26 mm d^{-1} rate for the MOSES-PDM data.

Comparison of the daily calculated Hargreaves-Samani derived evapotranspiration with the 1999-2001 available MOSES-PDM inter-comparison data is useful to assess differences between the methods. The data in Figure 3.28 suggests that in general average catchment Hargreaves-Samani derived evapotranspiration displays a similar pattern to the MOSES-PDM derived data. The Hargreaves-Samani data show larger peaks in evapotranspiration during summer months, with approximately a 1.5 mm d^{-1} difference in the summer June-August peaks. The differences perhaps derive from MOSES-PDM data being generated for large grids across the UK, and thus more averaged temperature data used in the calculation that may not be representative across the Tern catchment as a whole. Despite this, a strong R^2 of 0.82 is found between the 1999-2001 data, Figure 3.29, indicating a good relationship between the data sets.

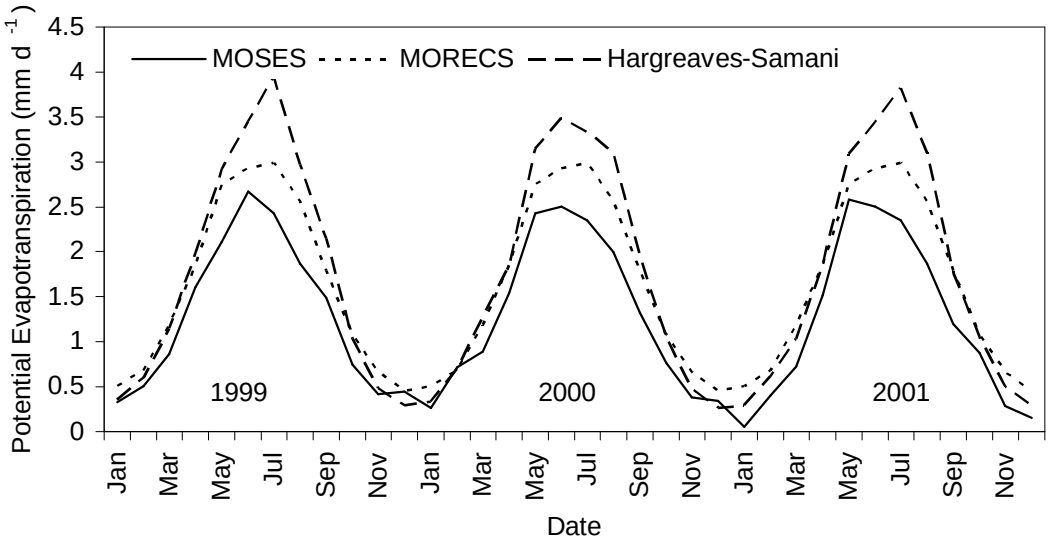


Figure 3.27. Comparison of monthly MORECS, MOSES-PDM and Hargreaves-Samani Eto

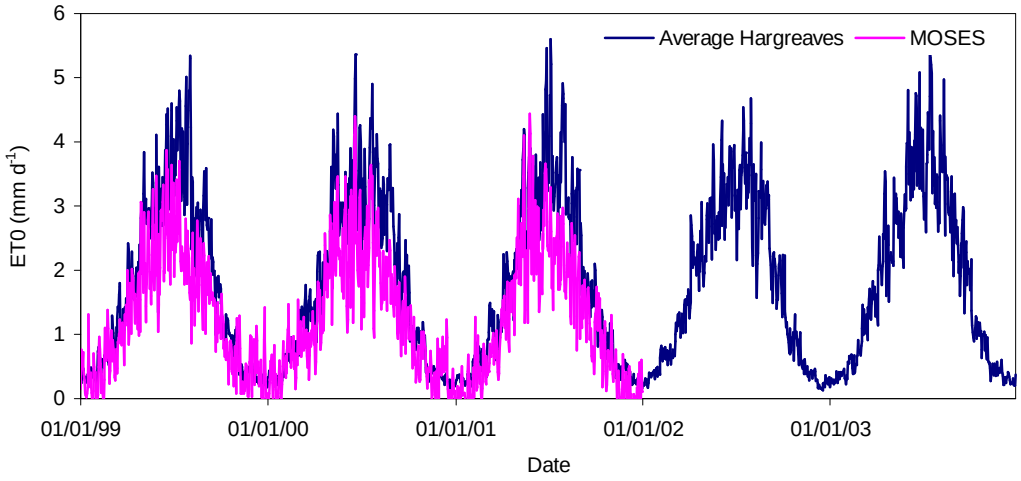


Figure 3.28. Comparison of calculated Hargreaves-Samani and modelled MOSES-PDM data

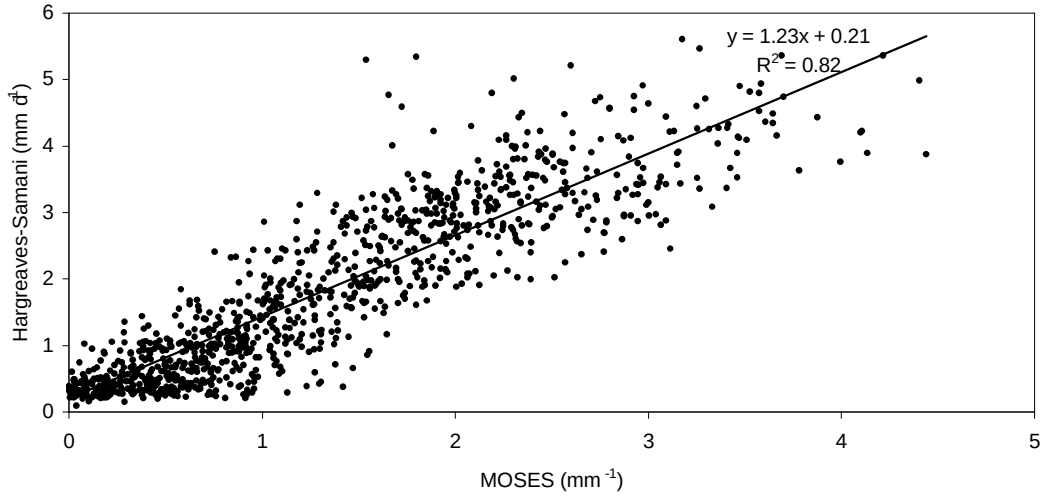


Figure 3.29. Relationship between Hargraeves-Samani derived Eto and MOSES-PDM Eto (1999-2001)

3.7. The Tern network and river flows

This section is divided into two main parts. Section 3.8.1 describes the main river network, including a description of the River Tern from source to mouth with additional details provided for the two main tributaries, the River Roden and River Meese. Section 3.8.2 follows to provide an analysis of the river flow data. In Section 3.8.2.1 the eight gauging stations to be analysed within the network and from different tributaries are presented. Summary statistics and analysis of flow are provided and individual gauge station reviews included so as to better understand the nature and errors from the observed data. Section 3.8.2.2 includes the presentation of the daily flow data, monthly flow analysis, flow duration curves and rainfall runoff relationships.

3.7.1. The Tern River network and main tributaries

The orientation of flow in the River Tern is predominantly from North to South. The source of the River Tern is in the grounds of Maer Hall in the North-East of the catchment. After flowing west through 'The Bogs' area of wetland, (Figure 3.30) the river passes through the first settlement of Norton-in-Hales. In the upper Tern, the river flows through ornamental lakes such as in Oakley Hall as well as various mill-ponds as it passes through Market Drayton. Coal Brook is the first main tributary to join the main river, just south of the town.

From the confluence of Bailey Brook with the Tern at Ternhill (Figure 3.30), the river then flows southward for the rest of its course. This middle section of the river is characterised by a large floodplain and has more confluences with other tributaries, notably the River Meese (Figure 3.30), and Potford and Platt Brooks which have been well researched due to influences of groundwater abstraction from the Shropshire Groundwater Scheme (SGS) (Mackay et al., 2006; Shepley et al., 2009).

The downstream section of the river is dominated by agro-industry, with notable abstractions and discharges from and to the river at the 'Dairy Crest' Dairy at Crudgington. The Rivers Strine and Roden adjoin the Tern in the lower section – with the River Roden confluence just upstream of the outlet gauge at Walcot (Figure 3.31).

There is a large storage area upstream of the gauge at Walcot (Figure 3.31) where water from the River Tern and tributaries is split to pass through three now non-operational sluice gates and a fish pass. As compared in Figure 3.32, this section of the Tern is particularly liable to flooding. Downstream of Walcot gauge (Figure 3.30) the river flows the remaining 6.4 km, passing through the National Trust's Attingham Park estate at Atcham, before draining into the River Severn (Figure 3.30).

The River Roden is the largest tributary of the Tern (259km^2) with a mean long term flow of 1.92 m^{3}s^{-1} at Rodington (see Figure 3.33 for location). The Roden flows in a north-west to south-east direction, passing through the small towns of Wem and Shawbury. The headwaters and upper half of the catchment flow over mudstones and as a result has a higher runoff component than in the rest of the catchment.

The River Meese is the second largest tributary to the Tern, with a drainage area of approximately 170km^2 and long term mean flow of 1.15 m^{3}s^{-1}. The river flows over Permo-Triassic Sandstones in the east of the River Tern catchment, and is characterised with a high baseflow component. As a result, flows are influenced by the Shropshire Groundwater Scheme (SGS) when in operation (more details of this are provided in Section 3.9.2). The River Meese drains from Aqualate Mere, the largest natural lake in the West Midlands. The tributary of Lonco Brook converges with the river, which drains from the north before following a predominantly western path through agricultural land, before its confluence with the Tern just downstream of Potford and Platt Brooks in the middle reaches of the Tern (Figure 3.30).

3.7.2. River flow

This section continues the discussion of the Tern river network by providing analysis of the gauged flow data at eight locations within the catchment. Included in Section 3.8.2.1 are summary statistics and meta data for each gauge. In Section 3.8.2.2 river flow analysis includes presentation and description of the daily flow data, monthly flow analysis, flow duration curves and rainfall runoff relationships.

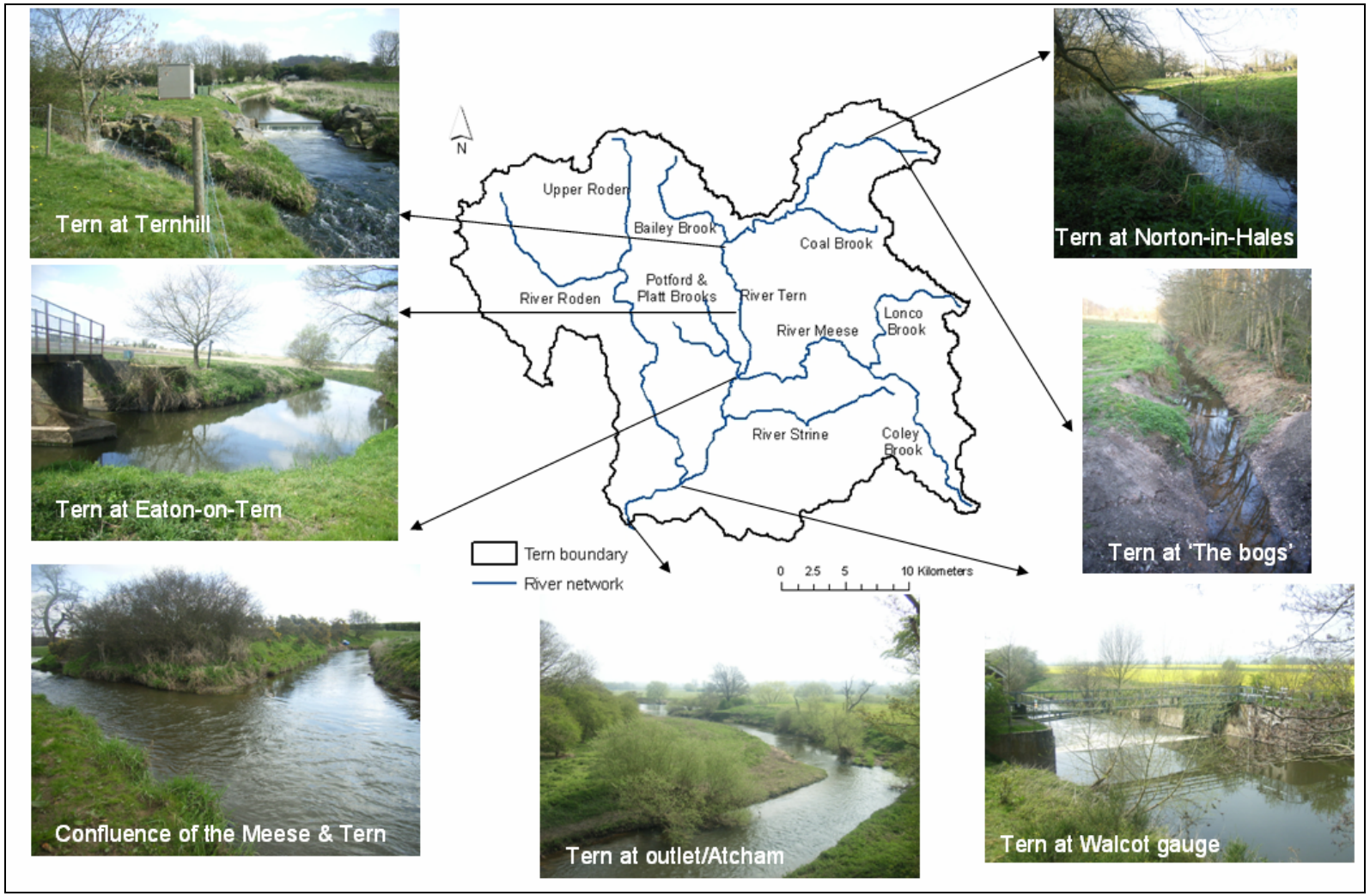


Figure 3.30. Key locations along the course of the River Tern

Figure 3.31. Storage area at Walcot, Photo by R. Rochester 4/4/2007. The confluence of the Roden with the Tern is shown in the foreground

Figure 3.32. Tern at Walcot, in Flood. Photo by Christopher Bayley (BBC news website) 12/11/2005. The area shown is the same as above in Figure 3.31

3.7.2.1. Gauge descriptions and general statistics

Daily flow data have been acquired for Environment Agency monitored flow gauges, through CEH Wallingford as part of the LOCAR data holdings. The data have been quality controlled by the Agency and are available for eight gauges as shown in Figure 3.33. The location map indicates that the distribution of gauges within the catchment is good with many of the major tributaries monitored at the downstream points before their confluence with the Tern. There is however only one active gauge on the River Roden, the largest tributary within the basin.

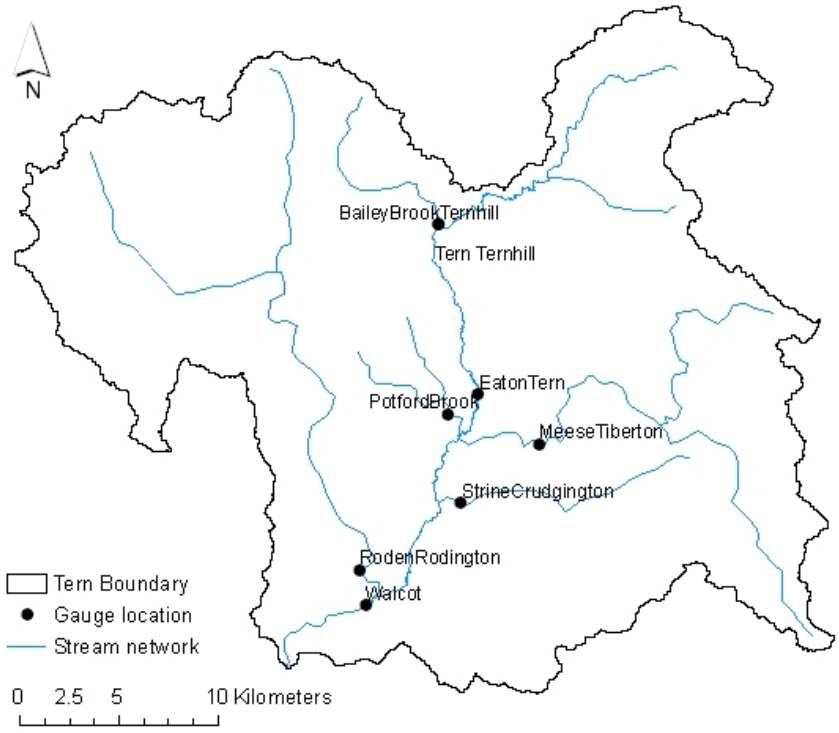


Figure 3.33. Locations of river flow gauging stations in the Tern Catchment

On the main Tern channel, there are three gauges; at Ternhill, Eaton-on-Tern and near the outlet, at Walcot. These gauges are well positioned to gauge upstream flows, flow downstream from the confluence of the River Meese, and at Walcot after all main tributaries have joined the network.

Table 3.21 summarises the key gauge statistics for each gauge and includes drainage area, Q95 and Q10, mean flow, stream length and S1085, the channel slope index. The gauges are shown to include a wide variety of drainage areas within the catchment, with both upstream and small tributaries as well as gauges at principal tributary outlets. Potford Brook and Bailey Brook (25km^2 and 34.4km^2 respectively) are the two smallest gauged sub-catchments, characterised with the smallest mean flows of 0.13 and 0.42 m3s^{-1}, respectively.

Generally, the ranges of mean flows vary in accordance with basin size with the exception of the Strine. The catchment is relatively large, 134km^2, but the mean flow is quite small, 0.67 m^{3}s^{-1} and is of the same order of magnitude as flow in Bailey Brook with a catchment 100km^2 smaller. The relatively small mean flows are perhaps due to the low lying nature of the Strine tributary, with a channel slope index of only 13.6, the smallest of all the streams. In addition, the flow in the catchment is altered by agricultural and industrial abstractions, with many irrigation and pipe drain channels. Flow is therefore likely to have been reduced in comparison to natural flow.

At the catchment outlet, Tern at Walcot, the mean catchment flow characteristics can be described. The mean flow of the River Tern is 6.83 m^{3}s^{-1} with Q95 flow of 2.32 m^{3}s^{-1} and Q10 flow of 13.45 m^{3}s^{-1}. The total upstream length has been defined as the length between Walcot gauge downstream and Maer lake upstream, and is 44.6km long. The channel slope index, S1085, is 61.1m, the largest slope index within the catchment although still relatively small considering the stream length.

Meta data for each flow gauge are presented in Tables 3.22 – 3.29 and give gauging station information including station descriptions, influences of gauged flow, photographs, flow duration curves, and mean annual flow analyses. These tables have been compiled to aid in a better understanding of the typical flows and quality of the historical records.

The gauging stations vary in the type of weir used to calculate discharge. Standard crump and rectangular weirs are located at Eaton-on-Tern, Meese at Tibberton, Bailey Brook and Tern at Ternhill. In addition, a non standard experimental design with trapezoidal flume and broad-crest weir is used to calculate discharge on the Roden at

Rodington. An electro-magnetic gauge using a bubbler device for water level measurement is situated within the Strine tributary. Although the weirs used to calculate flow vary, they have been significantly tested with stage-discharge relationships derived under a variety of water levels to ensure reliability of the data.

Table 3.21. Summary statistics for the eight primary river flow gauging stations

River name	**Gauge location**	**Location**	**Area (km^2)**	**Mean flow (m^3s^{-1})**	**Q95 (m^3s^{-1})**	**Q10 (m^3s^{-1})**	**Length up-stream of gauge (km)**	**S1085 (m)**
Strine	Crudgington		134.0	0.67	0.218	1.316	15.5	13.6
Potford Brook	Sandyford Bridge		25.0	0.13	0.045	0.223	7.0	15.4
Bailey Brook	Ternhill		34.4	0.42	0.112	0.857	11.8	27.1
Meese	Tibberton		167.8	1.15	0.454	2.036	26.2	31.4
Roden	Rodington		259.0	1.92	0.413	4.193	31.4	25.0
Tern	Ternhill		92.6	0.86	0.409	1.389	21.6	41.2
Tern	Eaton-on-Tern		195.6	1.67	0.68	2.936	30.8	49.0
Tern	Walcot		851.9	6.83	2.324	13.450	44.6	61.1

Table 3.22. Profile of Eaton-on-Tern river flow gauging station

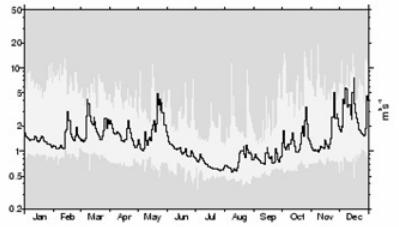
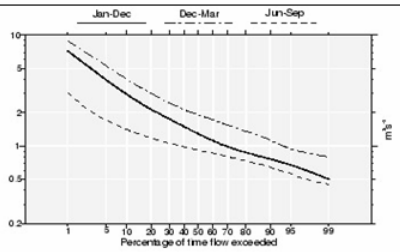
Station name:	**Tern at Eaton-on-Tern**	statistics of mean, max and min daily flows
Description	Two-bay Crump profile weir with identical crest heights, 6m total width, with crest tapping set into old mill sluices.	
NGR	33 (SJ) 649 230	
Quality		
Influences	Significant GW abstractions. Part of Shropshire Groundwater Scheme network. Agricultural and PWS abstractions balance effluent returns.	
Operation period	1972-2006	
		
Flow duration curve		gauge photograph

Table 3.23. Profile of Bailey Brook at Ternhill river flow gauging station

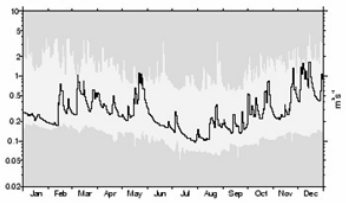
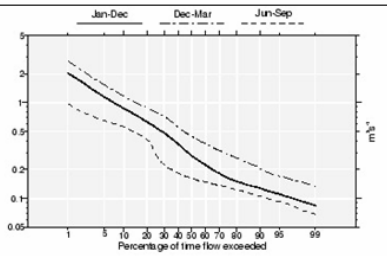
Station name:	**Bailey Brook at Ternhill**	Statistics of mean, max and min daily flows
Description	Rectangular notch, 1m wide, 0.63m deep with side contractions that acts as a broad-crested weir above notch full. Shares a recorder hut with the gauge (54044) Tern at Ternhill, the confluence being 10m downstream. Baseflow dominated regime but can be responsive. Small, low relief catchment instrumented for monitoring the Shropshire Groundwater Scheme. Solid geology: argillaceous rocks of the Lower Lias and Mercia Mudstones. Extensively overlain by Boulder Clay and glacial sand and gravel drift deposits.	
NGR	SJ 629 316	
Quality	Affected by irregular discharges and storage effects of Sandford Hall Lake	
Influences	Runoff influenced by groundwater abstraction/recharge. Runoff increased by effluent returns.	
Operation	1970-2006	
		
Flow duration curve		gauge photograph

Table 3.24. Profile of Roden at Rodington river flow gauging station

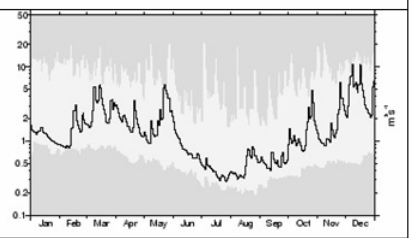
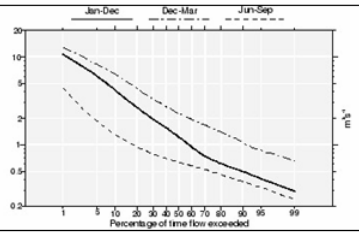
Station name:	**Roden at Rodington**	Statistics of mean, max and min daily flows
Description	Compound, non-standard structure consisting of a model tested trapezoidal flume and flanking broad-crested weirs within vertical sidewalls 7.3m apart. Built as an experimental model site for research purposes but has been retained as a flow measurement structure. Unresponsive rural catchment of gentle relief, underlain by sandstones, marls and clays of Carboniferous through to Liassic age, with extensive boulder clays and sand and gravel drift deposits.	
NGR	SJ 58901416	
Quality	Channel prone to weed growth. Weir drowns during high flows	
Influences	Abstractions balance effluent returns. Minor seasonal influence from spray irrigation. Shropshire Groundwater Scheme augments low flows.	
Operation period	1961-2006	
		
Flow duration curve		gauge photograph

Table 3.25. Profile of Strine at Crudgington river flow gauging station

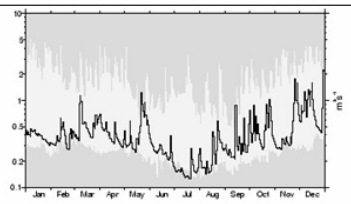
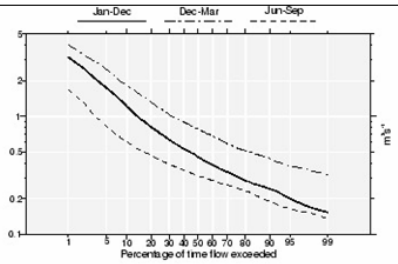
Station name:	**Strine at Crudgington**	Statistics of mean, max and min daily flows
Description	Electromagnetic gauge, using a bubbler device for level measurement, in trapezoidal channel. Very low velocities experienced. Current gauging replaced poor weed-affected open channel site in 1981. Early record not available. Very flat catchment draining Weald Moors. Headwaters dominated by argillaceous Carboniferous rocks. Lower reaches mostly Permian Sandstone overlain by mixed superficial drift deposits. Urban areas include Newport and the northern part of Telford.	
NGR	SJ 640 175	
Quality		
Influences	Runoff increased by effluent returns. Runoff reduced by industrial/agricultural abstraction.	
Operation	1982-2006	
		
Flow duration curve		gauge photograph

Table 3.26. Profile of Meese at Tibberton river flow gauging station

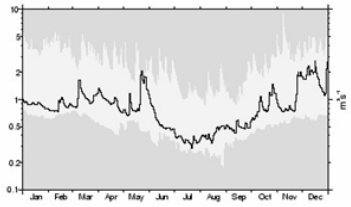
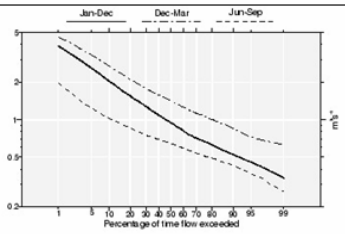
Station name:	**Meese at Tibberton**	Statistics of mean, max and min daily flows
Description	Crump profile weir, 6m wide, installed for the Shropshire Groundwater Scheme. Agricultural catchment of very low relief, with high baseflow component. Drains Sherwood Sandstone outcrop; intermittent Boulder Clay and glacial sand and gravel drift deposits.	
NGR	33 (SJ) 680 205	
Quality	Weir drowns only at very high flows, but channel recently re-graded to further reduce the risk.	
Influences	Runoff increased by effluent returns and reduced by industrial/agricultural abstraction.	
Operation	1973-2006	
		
Flow duration curve		gauge photograph

Table 3.27. Profile of Tern at Walcot river flow gauging station

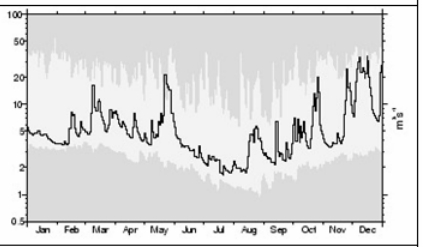
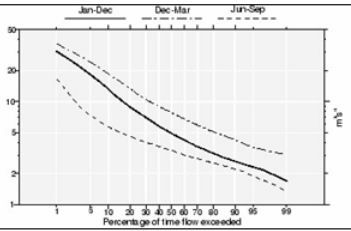
Station name:	**Tern at Walcot**	Statistics of mean, max and min daily flows
Description	Flat V weir, 15m wide. Bypass channel can be drained to enable work on weir, producing low or zero flows. Existing cableway. Siltation occurs upstream and downstream of the weir structure.	
NGR	SJ 59121227	
Quality	Prone to weed growth leading to an unstable stage-discharge relationship. Weir drowns out at high flows	
Influences	Regional groundwater pumping for Severn regulation. Industrial effluent from Wellington and Newport; abstractions for spray irrigation.	
Operation	1960-2006	
		
Flow duration curve		gauge photograph

Table 3.28. Profile of Potford Brook at Sandyford Bridge river flow gauging station

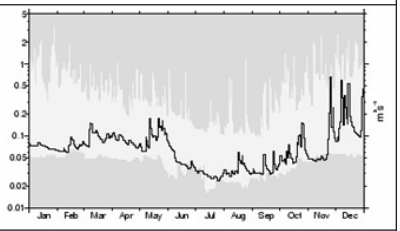
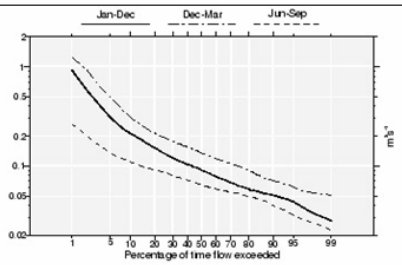
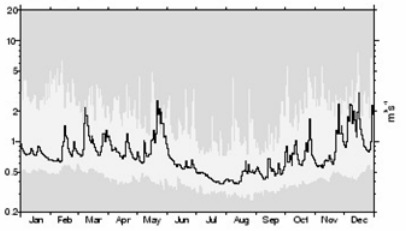
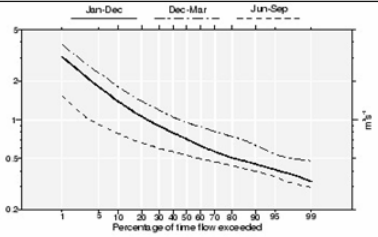
Station name:	**Potford Brook at Sandyford Bridge**	Statistics of mean, max and min daily flows
Description	Prefabricated flat V Crump profile weir, initially installed for the Shropshire Groundwater investigations. Site closed 1979-86 and repositioned in 1987 to avoid backing up (stated in National River Flow Archive). Reopened November 1986. Unresponsive, flat catchment on mixed geology: Sherwood Sandstone with Boulder Clay and glacial sands and gravel drift deposits.	
NGR	SJ 634 220	
Quality	Drowns at low flows due to seasonal weed growth. Heavy siltation at weir structure. Problems gauging at high flows above 0.2 m stage	
Influences	Runoff influenced by groundwater abstraction/recharge.	
Operation	1972-2006	
		
Flow duration curve		gauge photograph

Table 3.29. Profile of Tern at Ternhill river flow gauging station

Station name:	**Tern at Ternhill**	Statistics of mean, max and min daily flows
Description	Rectangular notch 4m wide by 0.43m deep with side contractions. Cableway for high flows. Not yet out of bank (to end of 2006). Shares a recorder hut with the adjacent Bailey Brook gauge.	
NGR	SJ 629 316	
Quality		
Influences	Runoff increased by effluent returns. Runoff reduced by industrial/agricultural abstraction. Irregular operation of sluices from Oakleigh Park may affect natural flow.	
Operation	1972-2006	
		
Flow duration curve		gauge photograph

The quality and reliability of data at each gauge are affected by a variety of factors, with many the same for each gauge. Weed growth is a particular problem, not just within the Tern catchment, but is a large scale problem affecting the quality of flow data. Within the catchment, Tables 3.27, 3.24, 3.28 note it is a particular problem listed at Walcot, Roden at Rodington, and Potford Brook. Similarly to weed growth within weirs, siltation is noted as a further problem at Potford Brook gauge (Table 3.28), and in the photograph of the weir, it can be seen to have been recently cleaned. Measurement of discharge during particularly high flows is a further problem – notably at Walcot, Roden at Rodington, Meese at Tibberton and Potford Brook. At very high flows the weirs can drown out and flood, thus any stage-discharge relationship becomes questionable.

Aside of data quality issues, Tables 3.22 – 3.29 note influences upon natural river flow. As Section 3.9.2 later reviews, the Shropshire Groundwater Scheme abstractions have the largest influence on natural flow when in operation. Aside of this, storage and irregular release of water from storage lakes such as at Sandford Hall Lake (affecting Bailey Brook flows, Table 3.23) and Sluice operation at Oakleigh Park (affecting Tern at Ternhill flows, Table 3.29) are important to note within the catchment. Tables 3.22 – 3.29 also note that Public Water Supply (PWS) and agricultural/irrigation abstractions roughly balance effluent returns to the streams.

3.7.2.2. River flow analysis

The daily time series of data for each gauge are presented in Figure 3.34. It can be seen that the records are comprehensive from the 1970s onward. River flow at Walcot, the most downstream gauge in the catchment has been recorded from the 1960s onward. The Potford Brook record, and Strine at Crudgington records are the most fragmented, and data is unavailable for the Strine prior to 1981 when the current electro-magnetic gauging came into operation following poor quality measurements prior to this. The data show that flow in the catchment, although varying in magnitude, is of a similar pattern for all gauges. For example, the signal of large flow peaks in 1990 and 2000 are found at the same time in all gauges.

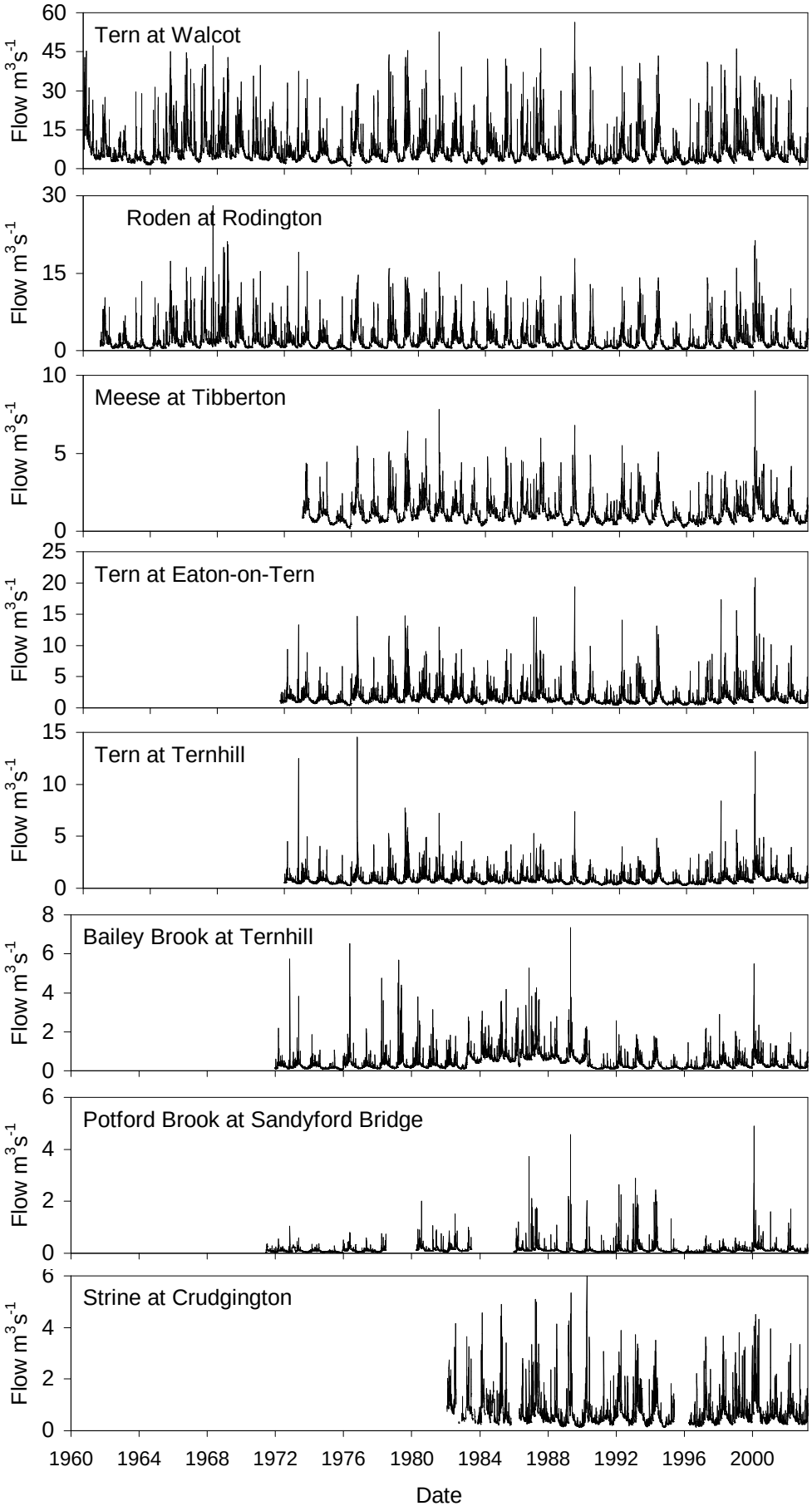


Figure 3.34. Presentation of daily river flow data

Long-term mean monthly flows for each gauge are displayed in Figure 3.35. The figure highlights the high and low flow seasons from December-February and July-September, respectively. Long-term January flows at Walcot average 11.28 m^3s^{-1}, this is approximately four times greater than low flows in July averaging 3.64 m^3s^{-1}. Figure 3.35 also highlights the same patterns in mean monthly flow for all the flow gauges in the catchment. The flows from Bailey Brook and Potford Brook are shown to be considerably smaller than for the other gauges, with flows not exceeding 1 m^3s^{-1} even during winter months.

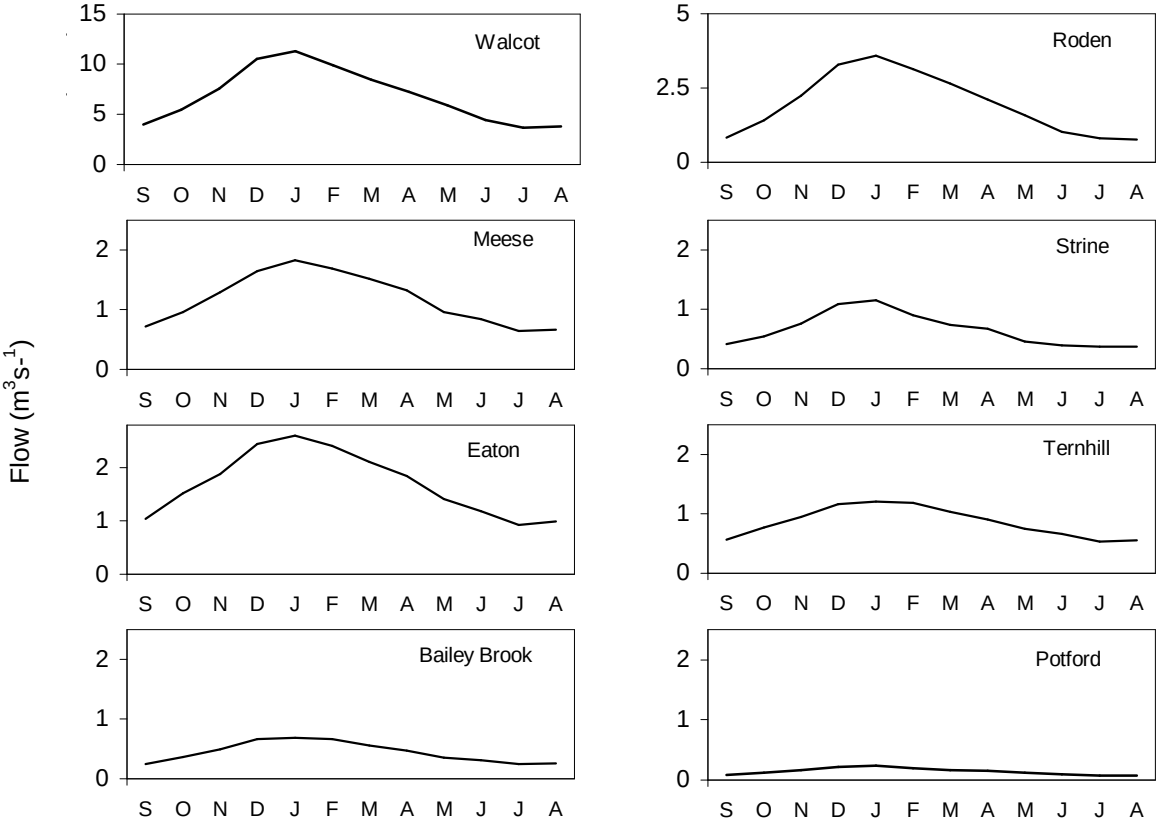


Figure 3.35. Mean monthly river flows for each gauging station
(note y axes for Walcot and Roden are a different scales, x axes are months September-August)

Further analyses of the flow data using flow duration curves (FDCs) to assess percentage of time that a given flow is exceeded (Shaw, 1994) have been calculated in Figures 3.36 and 3.37. Figure 3.26 displays the long-term flow duration curves (from the length of the available records). Aside of differing quantities between gauges, there are similar patterns for each curve with only the rivers Roden and Meese displaying any variance. Figure 3.37 furthers the FDC analysis by normalising flow for each gauge relative to the long term mean for the 1999-2003 modelling period to enable the FDCs comparison more easily.

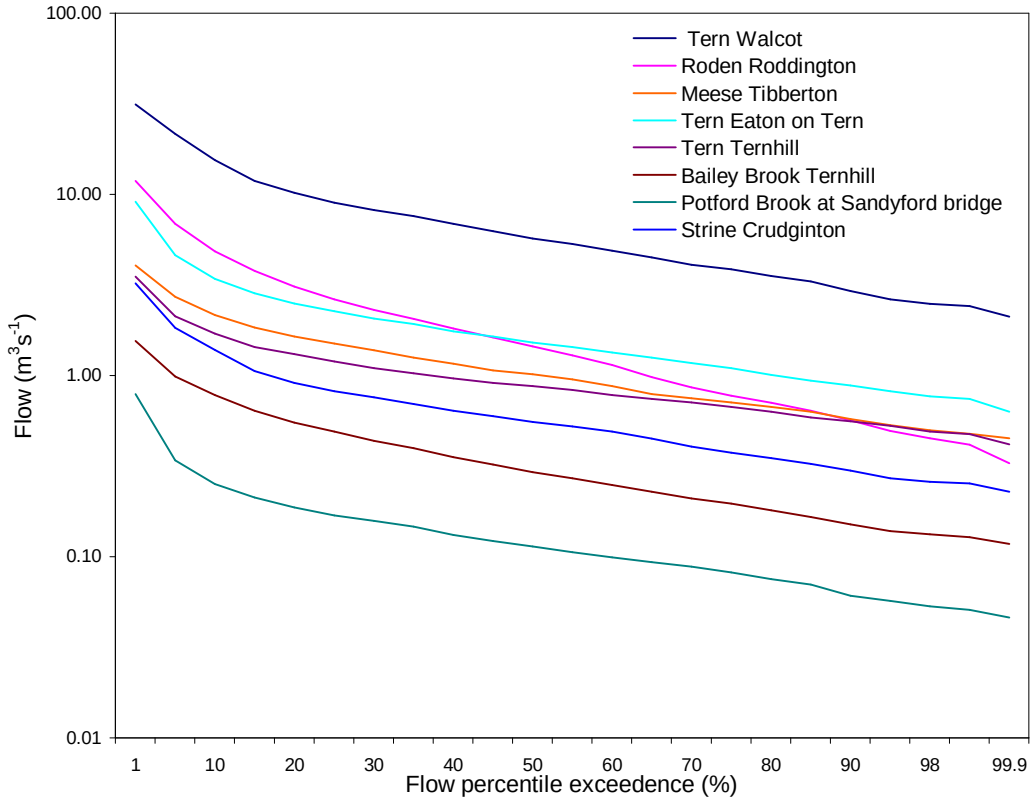


Figure 3.36. Long-term flow duration curves for tributaries within the Tern

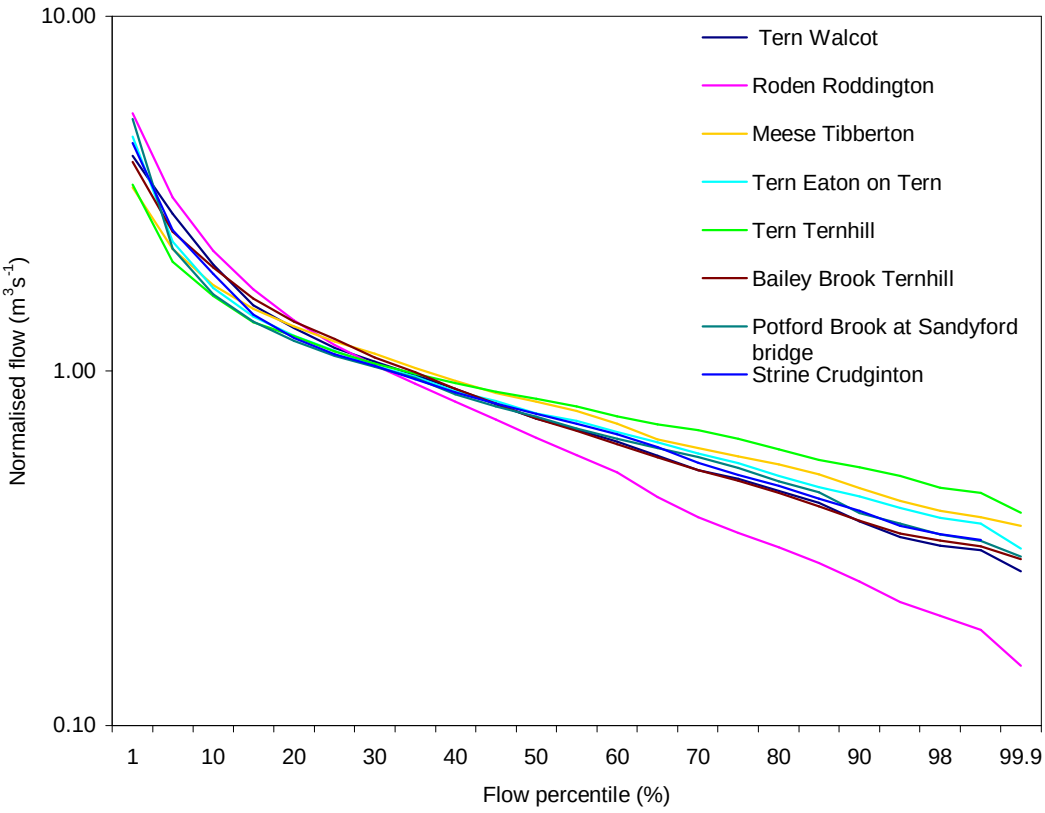


Figure 3.37. Normalised (flow/mean flow) flow duration curves (1999-2003) for tributaries within the Tern

Gustard et al., (1992) assessed 845 catchments from both permeable and impermeable substrates to derive 15 flow duration curves for catchments of different permeabilities. Results indicated that permeable catchments are expected to have a high Q95 and a low Q5 relative to their mean flow, which would result in a gentle FDC slope (Gustard et al., 1992). As Figure 3.37 shows, the gradients of the curves are relatively flat, indicating the nature of the permeable sandstone catchment the Tern is located in.

The Roden is shown as the fastest responding and flashier of the tributaries, likely a result of the headwaters and upper part of the catchment flowing over relatively impermeable mudstones, and can be seen from the respectively steeper sloping flow duration curve. The Meese and Tern at Ternhill tributaries are seen as having a slowest response with flatter flow duration curves, likely a result of the permeable sandstones that underlay these sub-catchments.

Figure 3.38 plots daily flow data from each gauge against Tern catchment mean daily precipitation from the eight rain-gauges used in the 1999-2003 analysis in Section 3.6.1.2. As intra-basin variation in precipitation is not that large within the catchment, as shown in Section 3.6.1.2, the average catchment rainfall has been used in the calculation. The results show the rainfall runoff relationship to be weak for every gauge with the best derived R^2 being 0.25. The relationship is not improved even when offsetting the daily flow gauge data by a day to the precipitation data, to account for lag time between the rainfall and runoff period, a method often used in rainfall-runoff analysis (Shaw, 1994).

The relationships shown in Figure 3.38 are likely to be explained by the porus sandstone bedrock beneath the catchment. In such a case, the precipitation falling within the catchment appears to be infiltrating the sub-surface and entering the river primarily as baseflow. The Meese at Tibberton rainfall-runoff is notably poor (R^2 = 0.09) and re-confirms the baseflow dominated nature of this tributary as shown in the flow duration curves above. Bailey Brook at Ternhill shows amongst the best rainfall-runoff relationship – likely the result of the bedrock in this area (Mercia mudstones) being more impermeable and resulting in a slightly better response of the streams to precipitation events. Although the Roden catchment is also dominated by mudstone, it is larger and therefore the rainfall runoff response may be slower, hence the slightly weaker relationship.

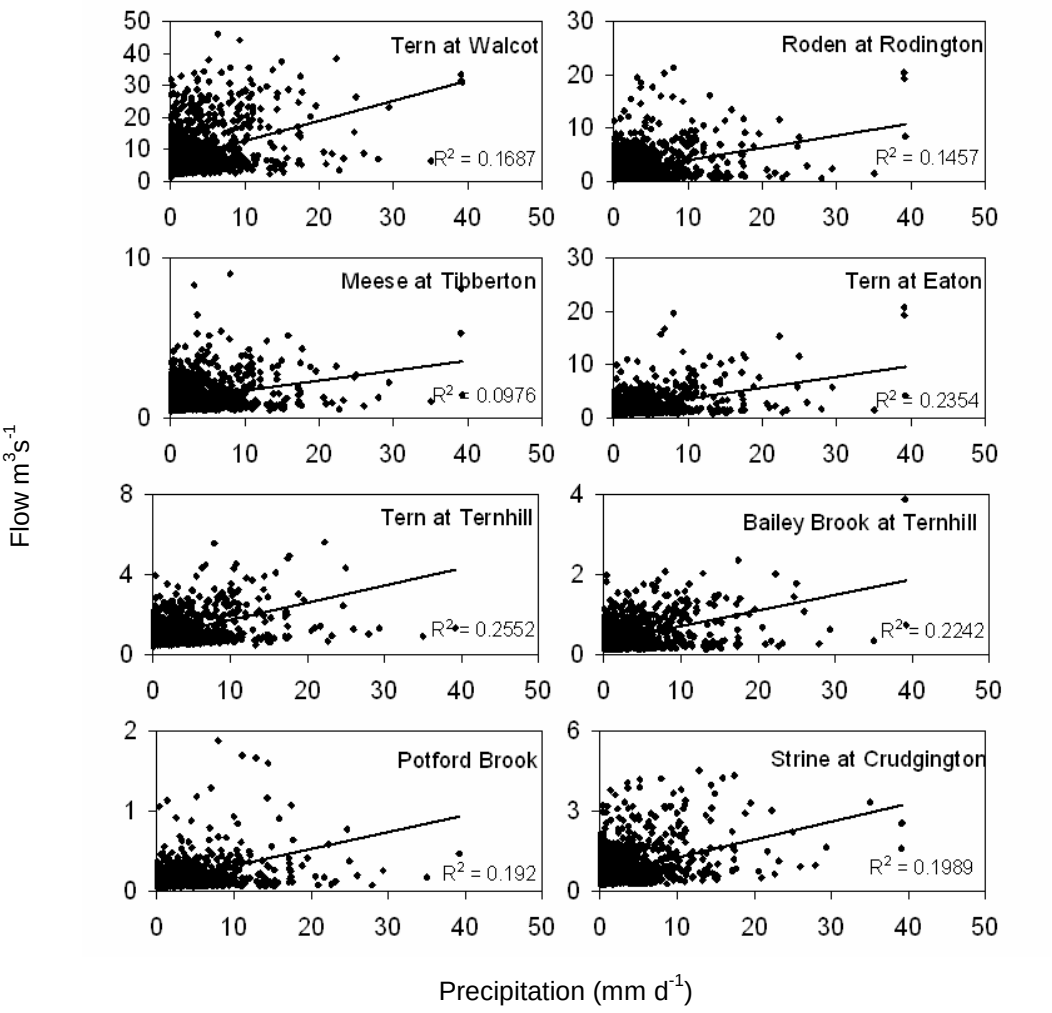


Figure 3.38. Rainfall runoff relationships for the gauging stations within the River Tern Catchment

3.8. Groundwater levels and abstraction

This section describes both groundwater level observation data from boreholes within the Tern catchment in Section 3.9.1, as well as details of groundwater abstraction as part of the Shropshire Groundwater Scheme (SGS) in Section 3.9.2.

3.8.1. Groundwater level data analysis

Divided into two parts, firstly the methodology for borehole site selection is given and an overview of the data provided. This includes the overall distribution and description of each site and the geology within which the borehole is located. The selection criteria is given, followed by descriptions of each of the records.

3.8.1.1. Borehole selection

Groundwater level data have been acquired from the Environment Agency (Midlands region) from borehole sites within the Tern catchment. In total, 36 records were available and a representative selection has been made based on the best spatial distribution, the best number of observations within the 1999-2003 period, and the best representation of different geologies; of both solid and drift deposits as well as drift thickness.

Figure 3.39 indicates the locations of 14 selected boreholes. The records are well distributed within the central part of the catchment. North-west and north-east of the catchment are lacking in available groundwater level records (notably because the EA focus is on monitoring the sandstone in the catchment as will be discussed in the subsequent section as part of the Shropshire Groundwater Scheme). The records are all of non-equidistant time steps, but approximate monthly recorded observations in each well. The exception is that Coley Farm record in the south-east of the catchment is of a high resolution hourly time step.

Table 3.30 summarises key features of each record, highlighting the three different types of solid geology in which the boreholes are located: two from Wildmoor and Wilmslow sandstone, five from Kidderminster and Chester Pebble Beds and seven from the Bridgnorth and Kinnerton Sandstone Formation. All three classes are grouped within the Sherwood Sandstone group as reviewed in Section 3.3. Table 3.30 also indicates the selected records are located within a range of differing drift deposits which are of different thickness. The records are located in a range of areas where there are either no superficial drifts, sands and gravels, diamicton, or on silty clay. The depths of the deposits range from 0 to 13.42m.

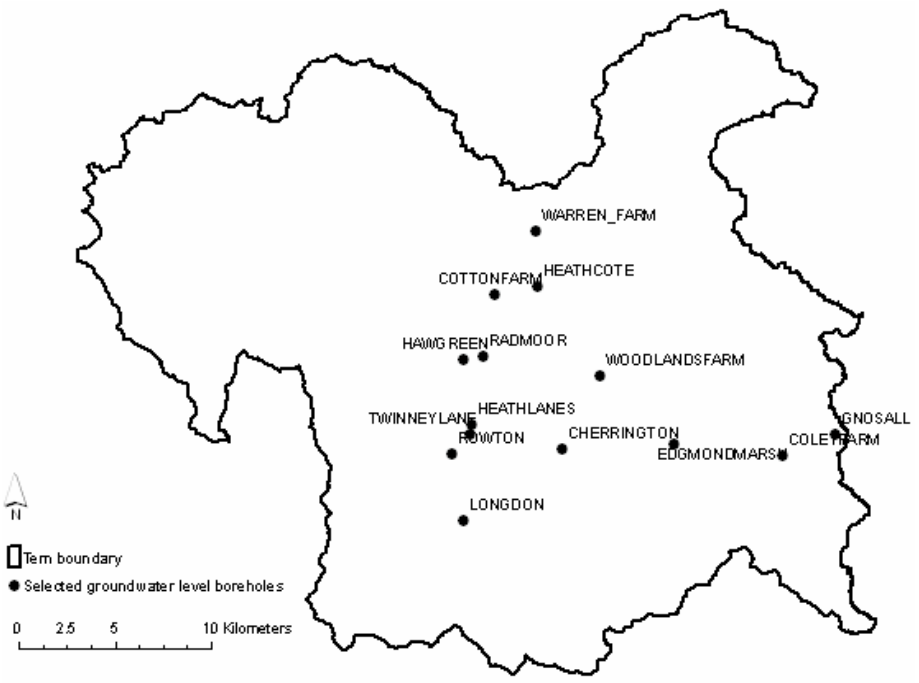


Figure 3.39. Groundwater level monitoring boreholes

Table 3.30. Selected groundwater level records and attribute information

Name	EA Code	E	N	No. obs	Geology Solid	Geology Drift	Drift thick (m)
COTTON FARM	2144GW	363150	327850	38	Bridgnorth/ Kinnerton	Silty clay	0.35
RADMOOR	0330GW	362570	324680	69	Bridgnorth/ Kinnerton	Diamicton	1.58
ROWTON	2062GW	360920	319480	21	Bridgnorth/ Kinnerton	Silty clay	1.72
HEATHCOTE	0168GW	365380	328340	41	Bridgnorth/ Kinnerton	Sands & Gravels	8.32
LONGDON	2058GW	361510	315970	44	Bridgnorth/ Kinnerton	Sands & Gravels	10.38
HEATHLANES	0169GW	361950	321050	54	Bridgnorth/ Kinnerton	NA	NA
HAWGREEN	2086GW	361500	324510	55	Bridgnorth/ Kinnerton	NA	NA
WARREN FARM	0410GW	365290	331250	44	Kidderminster/ Chester Pebble Beds	Sands & Gravels	13.42
WOODLANDS FARM	2107GW	368650	323620	44	Kidderminster/ Chester Pebble Beds	NA	NA
EDGMOND MARSH	0496GW	372480	320020	52	Kidderminster/ Chester Pebble Beds	Diamicton	NA
CHERRINGTON	0075GW	366690	319730	44	Kidderminster/ Chester Pebble Beds	Diamicton	NA
TWINNEY LANE	2081GW	361880	320560	27	Kidderminster/ Chester Pebble Beds	NA	NA
GNOSALL	1067GW	380860	320530	51	Wildmoor/ Wilmslow	NA	NA
COLEY FARM	26227GW	378130	319420	1000	Wildmoor/ Wilmslow	Sands & Gravels	9.18

3.8.1.2. Data presentation and analysis

The groundwater level data are presented in three groups according to the solid geology in which they are found. The Wildmoor and Wilmslow sandstone group, the Bridgnorth and Kinnerton sandstone group and the Kidderminster and Chester pebble beds.

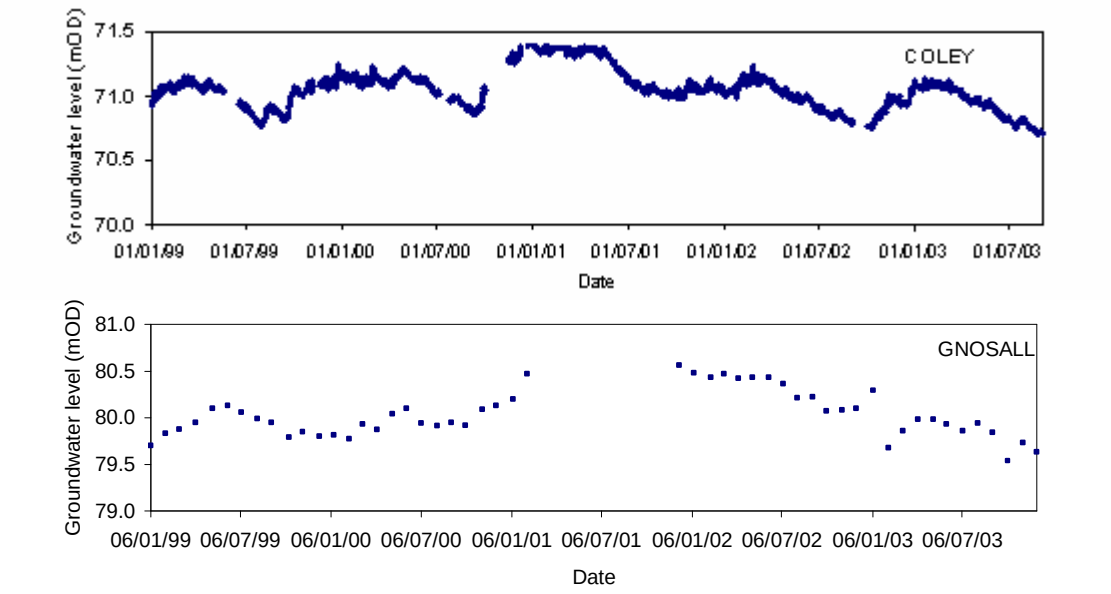


Figure 3.40. Groundwater level data from the Wildmoor and Wilmslow Sandstone Formation

Level data from the Wildmoor and Wilmslow Formation are from the Coley Farm and Gnosall sites. As indicated in Figure 3.40, both sites are located in the east of the Tern catchment. The level observations differ in temporal resolution, with available hourly data for Coley Farm (where high resolution level fluctuations are seen) and approximately monthly observations available from Gnosall. The Coley Farm time series is continuous with the exception of three main breaks in the record as highlighted in Figure 3.40. Comparatively, the Gnosall record is missing observation data for the most part of 2001. The two records are in good agreement and show similar patterns, with notable winter groundwater recharge, especially during January 2001. At Coley Farm (with the only high resolution time series) the annual decrease and recharge are shown to be in balance with no trends of gradually increasing or decreasing levels shown.

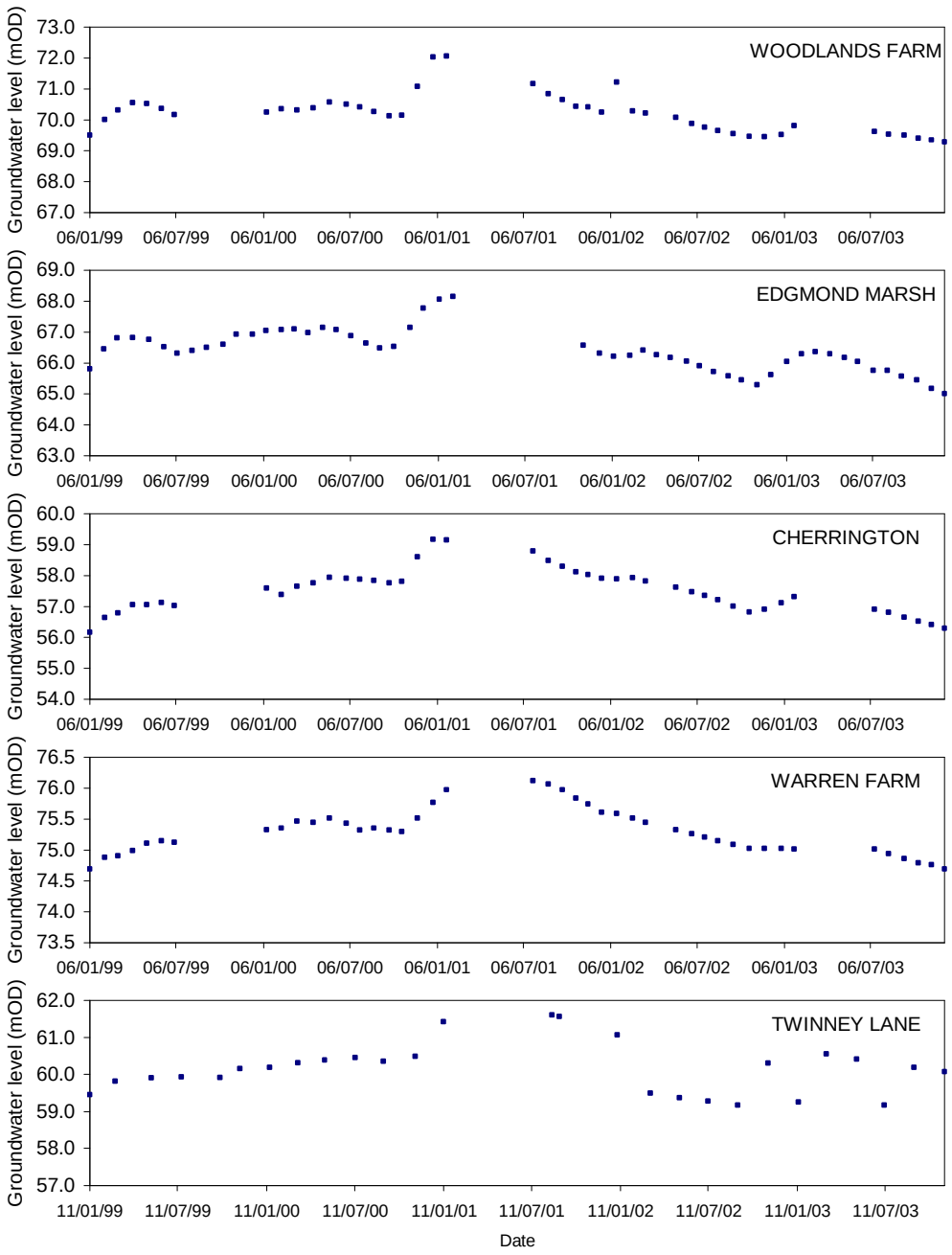


Figure 3.41. Groundwater level data from the Kidderminster and Chester Pebble Beds

The Kidderminster and Chester Pebble Bed groundwater level data are shown in Figure 3.41 for the five selected records. All of the records show very similar trends and agreement, with steadily fluctuating levels in 1999 and 2000 and a peak recharge event in January 2001 of ~1.5m. All records then show a gradual decline in groundwater levels of ~2m in 2002 before another smaller recharge event (especially notable in the Edgmond Marsh and Cherrington wells) in January 2003. All of the records are missing data during the spring – autumn 2001 period inclusive and as such makes detailed analysis of the data difficult.

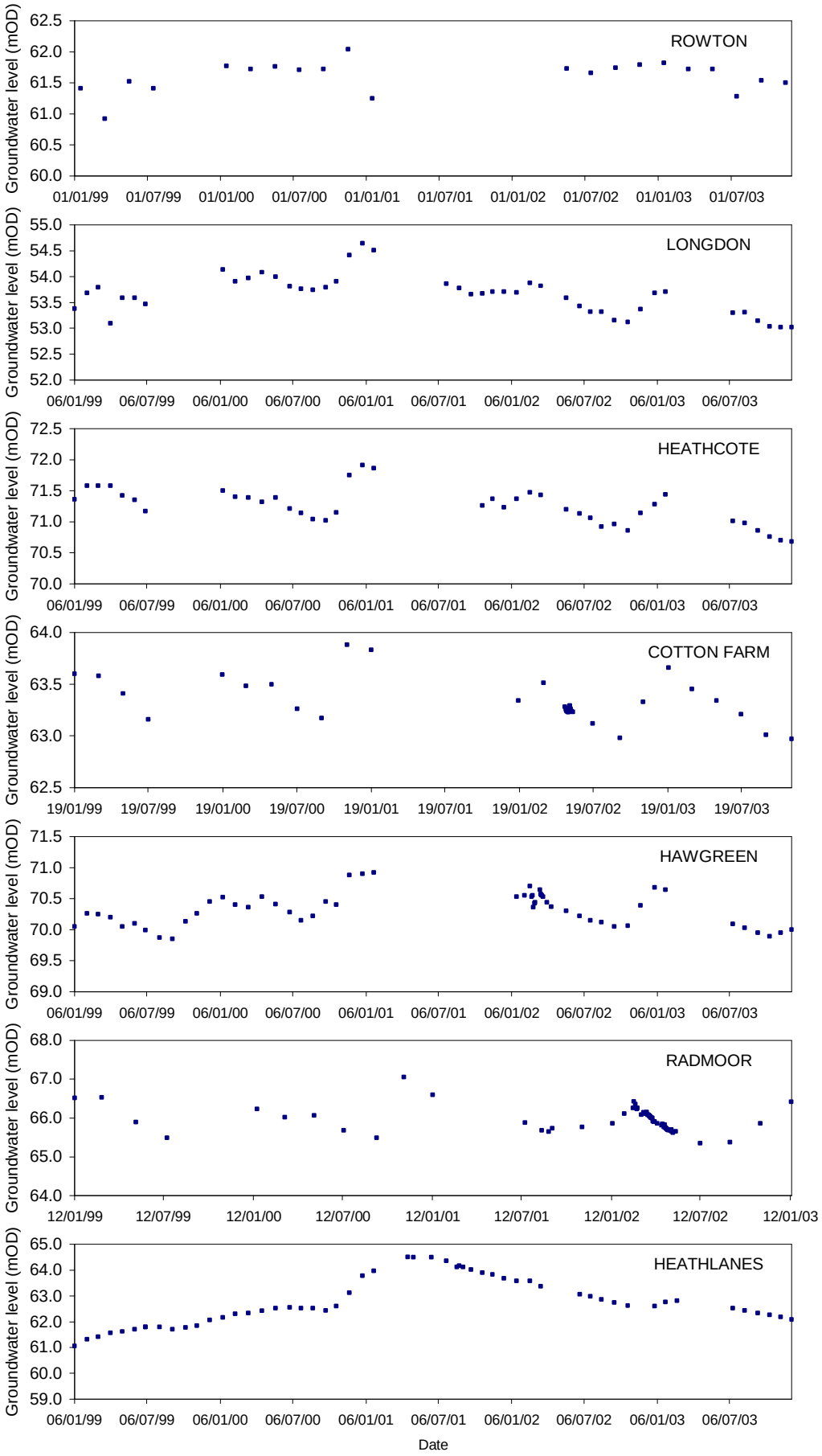


Figure 3.42. Groundwater level data from the Bridgnorth and Kinnerton Formation

The seven selected records from the Bridgnorth and Kinnerton Formation, shown in Figure 4.42, appear relatively fluctuating with more disparity between the individual records. Section 3.3.4 detailed that the hydraulic conductivity was most variable in this sandstone formation compared to the others in the catchment. This was due to the formation showing greater lithological diversity and therefore it is suggested that this may be contributing to the differences in the groundwater level records shown in Figure 4.42. The exception to this are the Heathcote and Longdon records that appear very similar and as highlighted in Table 3.30, are both from sands and gravels, although not within close proximity to each other. The Heathlanes record is the most complete with the fewest gaps in 2001, where all the other records are missing summer observations.

Table 3.31. Summary groundwater level statistics

Name	Solid Geology	Min level (mOD)	Max level (mOD)	Range (m)	Mean range (m)
COLEY FARM	Wildmoor / Wilmslow	70.7	71.4	0.7	0.9
GNOSALL		79.5	80.6	1.0	
COTTON FARM	Bridgnorth / Kinnerton	63.0	63.9	0.9	1.6
RADMOOR		65.4	67.1	1.7	
ROWTON		60.9	62.0	1.1	
HEATHCOTE		70.7	71.9	1.2	
LONGDON ON TERN		53.0	54.6	1.6	
HEATHLANES		61.1	64.5	3.4	
HAWGREEN		69.9	70.9	1.1	
WOODLANDS FARM	Kidderminster / Chester pebble beds	69.3	72.1	2.8	2.6
EDGMOND MARSH		65.0	68.2	3.1	
CHERRINGTON		56.2	59.2	3.0	
TWINNEY LANE		59.2	61.6	2.4	
WARREN FARM		74.7	76.1	1.4	

Further analysis and comparison of the level observations given in Table 3.31 indicate that the Wildmoor and Wilmslow Formation has typically the smallest range in water levels during the 1999-2003 period, and the Kidderminster/Chester Pebble Beds have the largest range in levels, a mean of 2.6m.

As previously introduced, The Tern catchment falls within the boundary of the Shropshire Groundwater Scheme (SGS). A brief introduction to groundwater abstraction in the Permo-Triassic Sandstones of the UK is included in the next section, and sets the context for the description of the SGS which details the purpose, usage and licensing of the scheme.

3.8.2. Groundwater abstraction in the Tern catchment

As highlighted in Figure 3.43, the Severn Trent region abstracts 409 million cubic metres of groundwater per year from its Permo-Triassic Sandstone is second only to the Thames region at 572 million cubic metres per year. The Permo-Triassic Sandstones of the UK are mostly contained within deep sedimentary basins in the west of England (Shropshire, Staffordshire and Cheshire) and on the east and west flanks of the Pennines.

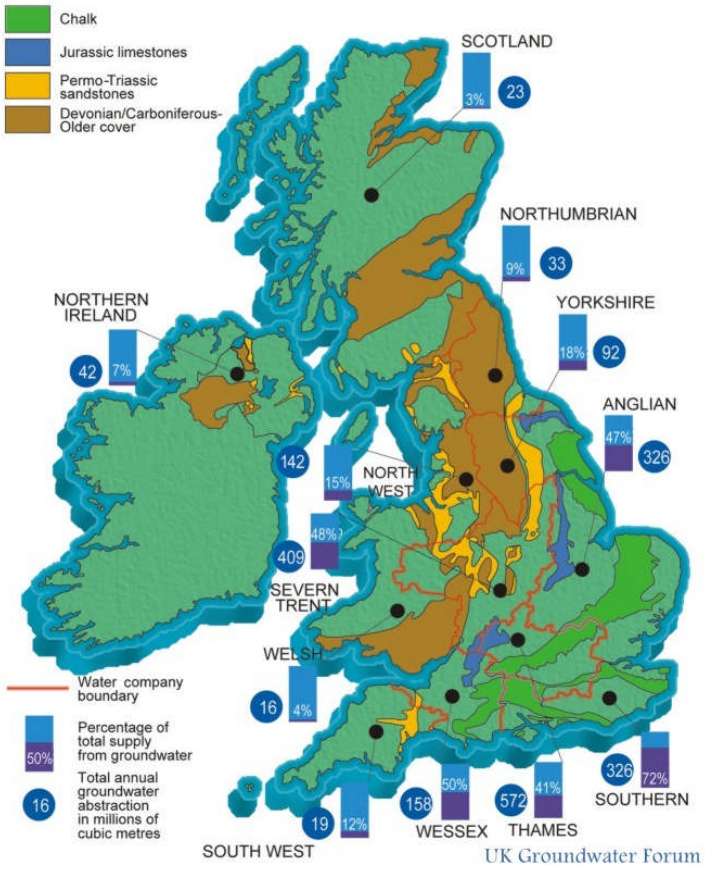


Figure 3.43. Groundwater supply in the UK (source: UK Groundwater Forum)

As already detailed in Section 3.3, the Permo-Triassic sandstones are a soft compact rock that is only weakly cemented enabling good flow through the individual grains of the rock's matrix, with the flow considerably enhanced along fractures and faults. It typically is very porus, with a porosity of ~30% and a high specific yield of 20-25%.

The Shropshire Groundwater Scheme (SGS) in an Environment Agency managed system where groundwater abstractions from the underlying sandstone aquifer within the Shropshire region are pumped to nearby tributaries such as the River Tern, and are

used to compliment and maintain flow in the River Severn during dry periods. In usual practice, the flow of the River Severn is augmented by two upstream Welsh reservoirs on the Rivers Vyrnwy and Clywedog. Approval for the scheme was passed in 1981, and construction began in 1982 within a framework of eight planned development stages. Stage five has been recently completed (Shepley et al., 2009) but stages six to eight are not yet deemed necessary. The initiative seeks only to develop each stage as and when they are needed (based on water resource demands).

Research such as that of Procter et al., (2006) from ADAS seek to assess impacts of the SGS on agriculture where zones of crop vulnerability to groundwater abstraction are highlighted based on the underlying geology. Surface water levels and quality are also monitored as well as surveys of fish and aquatic invertebrates to assess any adverse or beneficial impacts of the scheme.

The SGS is intended for use in short intervals and is not licensed for continual abstraction. It is for use only when the Clywedog reservoir cannot meet the water resource demands for the River Severn. The licence specifies that it can be used for maximum of 100 days per year or 250 days per five year period. The intended use is for once every three years, or two in every five years (Shepley et al., 2009). On average it has been used between five to fifteen weeks pumping per year with each phase of the scheme typically able to add 30 000 to 60 000 m^3 d^{-1} to river flow (Table 3.32). This equates to an increase in flow between 0.35 and 0.7 m^3 d^{-1} for each phase in operation.

Table 3.32. Statistics of the Shropshire Groundwater Scheme
(Source: Environment Agency, 2009)

Max. no. of development phases	8
Max. no. scheduled groundwater pumping stations	64
No. principal environmental observation boreholes	91
No. major outfalls discharging in excess of 10 000 m^3 d^{-1}	9
No. minor outfalls discharging less than 10 000 m^3 d^{-1}	12
Max. design yield of scheme (max increase in net abstraction)	225 000 m^3 d^{-1}
Total length of pipeline	73km
Pipeline diameter ranges	250 – 700mm
Typical depth range of abstraction boreholes	60 – 200m
Range of abstraction borehole diameters	610 – 760 mm
Typical abstraction borehole yield	4000 – 7000 m^3 d^{-1}
Typical discharge per phase	30 000 – 60 000 m^3 d^{-1}

Within the Tern catchment, only phases one and four are in operation and located within the catchment, although in the future the River Tern will play a larger role in delivering abstracted water to the River Severn as phases five, six (associated with the River Roden) and eight (River Tern) are also located within the area. To date the SGS has been used in 1984, 1985, 1989, 1995, 1996 and 2006.

For the purpose of hydrological modelling in the following chapter, it is therefore notable to construct and develop a model for time periods in which the scheme has not been in operation (as more detailed discharge data and volumes are not available). For this reason the 1999-2003 period is a good choice.

3.9. Water balance of the Tern catchment

The water balance is a useful tool to describe the main flows of water in and out of a catchment (Shaw, 1994), where the volume of water inflow should balance with water outflow assuming no change in storage. Although the number of components used to express the balance can become quite large, the major terms are usually the outflow or discharge from the river network (Q), the input from precipitation (P), the loss from evapotranspiration (ET) (in reality determined by Actual Evapotranspiration (AET) rather than Potential Evapo-transpiration (PET) which is independent of water availability since it assumes that there is no water restriction on evapotranspiration), the inflow or outflow from the aquifer not defined within the limits of the catchment (G) and the change in storage or groundwater level (ΔS).

The water balance for the Tern catchment is shown for individual years from 2000 – 2003 in Table 3.33, with the ID listed for each part of the water balance to aid description. The groundwater movement and change in storage contributions are listed as '-' as the groundwater contributions (G) component is unknown for the 2000-2003 period. As the catchment is situated on predominantly sandstone aquifers that are not necessarily limited to the defined catchment boundary, there may be sub-surface movement across the catchment boundary that could be contributing, removing or balancing on a yearly basis. As the quantity of groundwater inflows and outflows are

not known, it has been necessary to assume that they are in balance, but it is noted as a limitation in the calculation. The change in storage or groundwater level (ΔS) on an annual basis was shown in the high temporal resolution record at Coley Farm (Figure 3.40) to generally be in balance with annual recharge during winter months increasing the groundwater level from the summer draw-down of a similar magnitude.

The water balance in the catchment is therefore defined in terms of [6] flow (Q) = Precipitation (P) – Evapotranspiration (though calculated with PET rather than AET and this therefore expected to result in the difference between the calculation shown in [6] and the observed river flow in [1]). The observed river flow data have been converted to the same unit (mm yr^{-1}) by dividing the total flow volume by the catchment area as P and ET so that the units are comparable in the calculation of the water balance.

The values of evapotranspiration [3] shown in Table 3.33 are those calculated by the PET Hargreaves-Samani method (Section 3.6.2), the same data as used later in the thesis as input to the hydrological models from which AET is later calculated as part of the simulations. The limitation of using PET rather than AET in these calculations is that evapotranspiration will be over-estimated in the calculation as PET is the maximum potential evapotranspiration from which AET is calculated based on soil moisture availability. The [2] precipitation and [1] river flow data shown in Table 3.33 originate from the data described in Sections 3.6.1 and 3.8 respectively, with these data being specific measurements within the catchment.

Table 3.33. Water balance in the Tern catchment (all units mm yr^{-1})

ID		**2000**	**2001**	**2002**	**2003**
[1]	River flow at basin outlet (Q)	357	298	247	194
[2]	Precipitation (P)	944	720	764	618
[3]	Hargreaves-Samani Potential Evapotranspiration (PET)	651	657	653	701
[4]	Groundwater (G)	-	-	-	-
[5]	Change Storage (ΔS)	-	-	-	-
[6]	**Q = P – PET** (Hargreaves-Samani)	**292**	**63**	**111**	**-83**

Note: Groundwater (G) and Change Storage (ΔS) are listed as ' – ' as data were unavailable for the groundwater movement across the catchment boundary

3.10. Summary

This chapter has provided a detailed review of the Tern catchment, and its associated hydro-meteorological, topographic, soil and geological data. A review of data type, extent and source is provided in Table 3.34. In addition, this chapter serves as a conceptual understanding of the different components within the hydrological system, as synthesised on a sub-catchment scale in Figure 3.44 where differences in key processes across the Tern catchment are shown. The description of the differences at the sub-basin scale facilitates the understanding of the different spatial variations internally within the catchment, as Figure 3.44 is described in the same terms (at the eight gauging stations) as for the later hydrological modelling and analysis of models undertaken in this thesis.

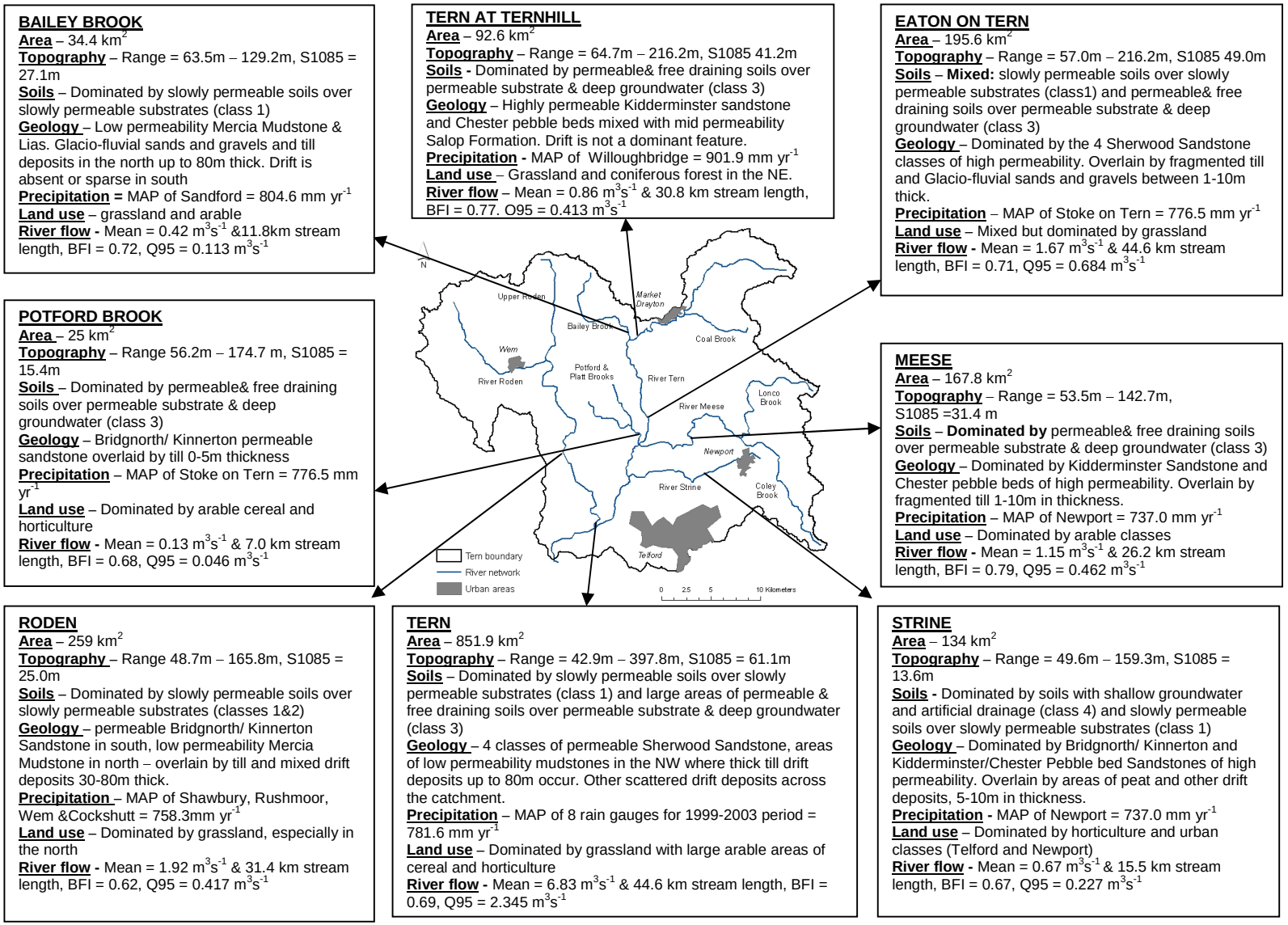
One of the main differences across the catchment shown in Figure 3.44 is the different geology in each sub-basin. This is a result of the large division in permeable sandstone classes (dominant in the Meese, the upstream areas of the Tern at Eaton and Ternhill, Potford Brook and the Strine) and low permeability mudstones across the catchment (dominant in the Roden and Bailey Brook). In addition to the mixed solid geologies, Figure 3.44 describes how each of the sub-catchments are characterised by different drift deposits overlaying the solid geology.

The conceptual model in Figure 3.44 also describes the differences in area, topography, soils (where the listed classes refer to the adopted five soil classes in Table 4.11), precipitation, land use and river flow at the same sub-catchment scale. The figure provides individual profiles for each of the different sub-basins and highlights the heterogeneous nature of the Tern catchment. The understanding acquired as a result of the data analysis in this chapter is used in the selection of an appropriate modelling code and construction of hydrological models in Chapter 4.

Table 3.34. Summary of data for the Tern catchment

Component	Data type / details / extent	Source
Topography	OS landform profile DTM 1:10,000.	Edina Digimap
Geology Solid	Solid geology type and extent.	British Geological Survey
Geology Drift	Drift geology type, spatial extent and thickness.	British Geological Survey
Geology Hydrological	Horizontal and vertical hydraulic conductivity and core porosity for different sandstones.	LOCAR Reports
Soils (Associations)	NATMAP vector data for catchment + 2km buffer zone, includes information on 32 soil associations + series from England and Wales Soil Classification.	Cranfield NSRI
Soils (Host)	1km gridded HOST for extent of Tern Catchment.	NERC via LOCAR / CEH Wallingford
Land use	LCM 2000 25m raster data of dominant vegetation cover.	NERC via CEH Wallingford
Precipitation	18 stations of gauged data, 7 used in the L-T analyses, 8 used in 1999-2003 modelling period analysis, daily data.	Environment Agency
Evapotranspiration	MOSES – Inter-comparison project data daily Eto between 1999-2001. MORECS – monthly averaged 1961 – 2006. Hargreaves-Samani calculated PET from temp min + max and incoming solar radiation 1991-2003.	Met Office British Atmospheric Data Centre (BADC)
Flow Network + Gauge Info	OS line data from Digimap, catchment spatial extent from BGS.	British Geological Survey (BGS)/ Ordnance Survey (OS)
Flow Discharge Data	8 gauges of daily flow data, 1960s onward.	NERC / NRFA
Groundwater Levels / Borehole Data	36 boreholes of which 14 chosen. Level data at sporadic intervals.	Environment Agency Boreholes

Figure 3.44. Conceptual model of the spatial variation of key processes at the sub-catchment scale in the Tern



Chapter 4

Development of homogenous & distributed River Tern hydrological catchment models

4.1. Introduction

One of the objectives of the research that was specified in Chapter 1 was to assess uncertainty in the different conceptualisations of hydrological models. This chapter describes the set up of two different MIKE SHE models with differing spatial representations. Section 4.2 details the component set-ups of the catchment models. The first model, a homogenous catchment model is described, where the majority of input data and model parameters are set to catchment representative uniform values. The second model, which is more spatially distributed, is also described simultaneously to enable model comparisons.

The general simulation specifications shared for both models are first discussed, followed by details of the topographic, precipitation, land use, evapotranspiration, overland flow, unsaturated and saturated zone flow representation. The models' conceptual parameterisations as well as the minimum and maximum feasible ranges for each parameter in the different components are also given within this chapter. The specific calibrated and defined parameter values are detailed further in Chapters 5 and 6.

Section 4.3 outlines the construction of a MIKE 11 hydraulic river model, including the network delineation and cross section specification. The same MIKE 11 model is used in both catchment models.

4.2. Component representation

Within the simulation specification the overland flow (finite difference), rivers and lakes (MIKE 11), unsaturated zone flow (2 layer UZ), evapotranspiration and saturated zone (finite difference) options have been included, as summarised in Table 4.1. As shown, the choice of solvers used for each hydrologic component is the same for both homogenous and distributed models. Although Section 2.3.3 suggested that some of the process representations and governing equations may not be applicable at large scales that represent the whole catchment, the solvers used in the models have been kept the same to enable more of a direct comparison of the different spatial representations of models in subsequent chapters.

Table 4.1. Summary of components included within the catchment models and the choice of solvers used for each component

Component	Tern homogenous model	Tern distributed model
Evapotranspiration	Christiaensen-Jensen method	Christiaensen-Jensen method
Overland flow (OLF)	Finite difference	Finite difference
Unsaturated zone flow	2 layer	2 layer
Saturated zone flow	3D finite difference	3D finite difference
River model	MIKE 11 hydraulic model	MIKE 11 hydraulic model

4.2.1. General simulation specifications

Both of the models are set to store results of river flow and groundwater levels, as well as components of the water balance on a daily time step between 01/06/1998 and 31/12/2003. As described in Section 3.6.1.2, this time period of five and a half years includes wide ranging variation in precipitation and has primarily been chosen as it is of interest to see how the models simulate the different wet (2000), average (2001) and dry (2003) years. The primary simulation years are 2000-2003, for which observed river flow and groundwater level data are available for which to calibrate and test the model. The model is set to begin simulation at the start of June 1998 which is prior to the main evaluation period to include an adequate warm up time for the model to stabilise from initial conditions (Stephenson and Freeze 1974, Refsgaard, 1997). Initial conditions of the models have been obtained using previously completed simulations for both homogenous and distributed set-ups, by use of a ‘hot start’ iterative process. The initial

conditions for starting parameters (such as initial surface water depth) are therefore more realistic, as they are obtained from the simulated values of these parameters from a date well into the simulation period, after which the models should have stabilised, in this case the end of July 2002, which is a representative summer period with similar conditions to that of the June 1998 start date.

Settings for the time step controls of the models are summarised in Table 4.2. These settings are important in fixing the maximum model time allowed to solve each of the model components listed in Table 4.1. Time step in the unsaturated flow (UZ), overland flow (OL) and evapotranspiration (ET) components are always the same and are set to 24 hours. The maximum UZ, OL, ET time step (even integer multiple of max SZ time-step allowed) are always less than or equal to the maximum SZ time step, which is also set to 24 hours.

Table 4.2. Settings of Time step control for different components of the Tern Catchment models

Time steps	**Value**	**Unit**
Initial time step	2	hrs
Max allowed UZ, OL, ET	24	hrs
Max allowed SZ	24	hrs
MIKE 11 time step	60	minutes
Increment of reduced time step length		
Increment rate (0-1)	0.5	
Parameters for precipitation dependent time step control		
Max precipitation depth per time step	10	mm
Max infiltration amount per time step	50	mm
Input precipitation rate requiring its own time step	0.1	mm hr^{-1}

The ‘increment rate’ is a factor for slowly increasing time step length back up to the maximum time step, after the time step has been reduced. This is important in allowing model stability and smooth simulations that do not abruptly jump back to set definitions after the time step has been reduced (DHI-WE, 2005).

The ‘maximum precipitation depth per time step’ and ‘maximum infiltration amount per time step’ are thresholds used when the amount of precipitation or infiltration in the current time step (i.e. 24 hours) exceed the specified amounts of 10mm and 50mm respectively. In such a case, the defined time step is then reduced until precipitation or infiltration return to within the allowed threshold. This improves numerical stability during rainfall events, but can lead to excessively small time steps during large storms.

The overland flow, unsaturated zone and saturated zone model components each have their own dialogues that enable specification of computational control parameters that control the way each component is solved. Default values are set within MIKE SHE when creating a new model and are usually adequate and need not be changed (DHI-WE 2005).

The catchment model domain has been defined using an ESRI Shapefile that codes area inside the catchment, the boundary of the catchment and area outside the boundary, as previously shown in Figure 3.1. In this way MIKE SHE is given information on how to pre-process the other input data that may cover a larger area than the defined catchment before model simulations begin.

The catchment is contained within a grid set within MIKE SHE, with the rectangle corner co-ordinates (specified to coincide with the British National Grid):

$X_{min} = 340000$ $X_{max} = 385000$

$Y_{min} = 300000$ $Y_{max} = 344000$

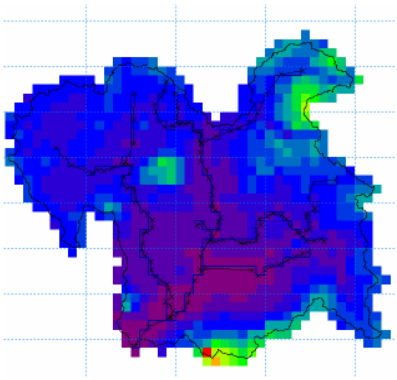
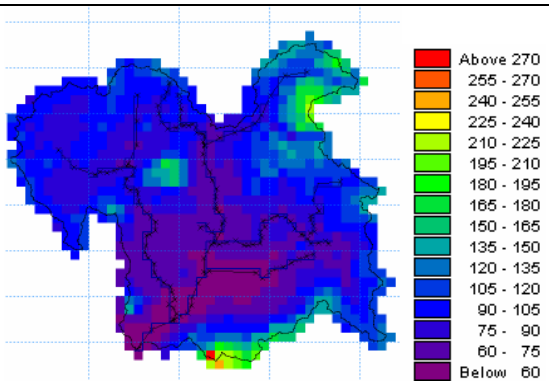
A grid size of 1 km x 1 km has been specified for the model. Using this grid size results in 1020 grid cells that are contained within the catchment. Having undertaken pilot tests to assess the optimal grid size, a 1000m grid was considered suitable in order to keep computational time to a realistic scale, whilst still being able to include spatial variation of input parameters within the distributed model. A typical model simulation running on a Pentium 4, 2.8 GHz completes within approximately six minutes for the homogenous model, and ten minutes for the distributed model.

4.2.2. Topographic representation

In MIKE SHE, the surface topography defines the drainage surface for overland flow, as well as the uppermost surface of both the unsaturated zone columns and the saturated zone model. In addition, the topography file acts as a point of reference from which to specify unsaturated zone and saturated zone parameters and observations, such as the depth of each layer and groundwater level observations.

Table 4.3 shows the topography representation is the same within the two models. The data are derived from the 25m DTM described in Section 3.2 and re-sampled to a 1 km × 1 km grid. Although the homogenous model is designed as a mostly uniform parameter model, as described, the movement of water within MIKE SHE is primarily governed by the surface topography and so it was considered important to also be spatially variable.

Table 4.3. Representation of topography in the two hydrological models

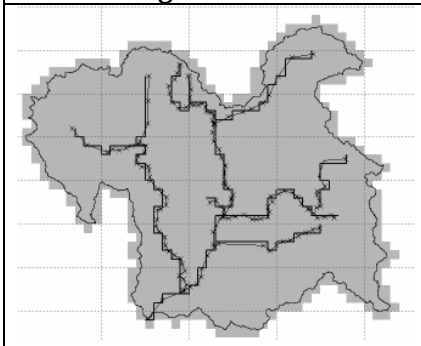
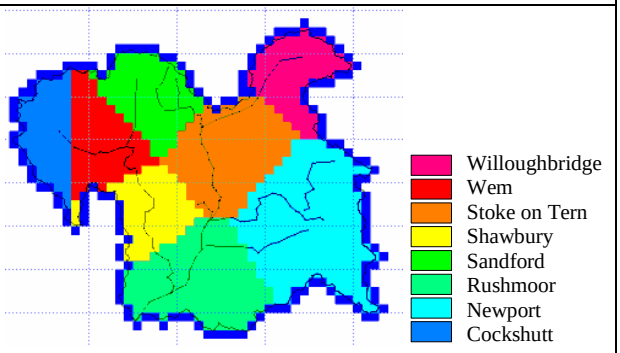
Tern Homogenous	**Tern Distributed**
	
Using the Profile DTM 1: 10, 000 elevation data as previously introduced (Section 3.2). The above plots display the processed data to the 1km × 1km grid. The data have been converted from raster to ASCII format, and developed to a MIKE SHE grid file in order to be compatible with the models. As shown, the same topography file is used for both the homogenous and distributed models.	

4.2.3. Precipitation

Table 4.4 summarises the differences in input precipitation between the homogenous and the distributed models. Catchment mean daily precipitation (mm) derived from the eight rain gauges described in Section 3.6.1.2 is used in the homogenous model and seeks to be representative of the whole catchment. Individual time series from the eight rain gauges comprising the catchment mean are used in the distributed model, which have been spatially defined using Thiessen polygons (Thiessen, 1911) within Arc GIS (Diskin, 1970).

Although Thiessen polygons can be considered as a semi-distributed approach of spatial representation, it was suitable to use this method as opposed to fully distributing the data by gridding it using a method such as Inverse Distance Weighting (IDW), that realistically require a higher sample of gauges than the eight used.

Table 4.4. Representation of precipitation in the two hydrological models

Tern homogenous	**Tern distributed**
	
The catchment areal mean precipitation of the eight primary rain gauges, introduced and analysed in Section 3.6.1.2, has been input to the model in a MIKE SHE time series file, as spatially uniform in distribution, with rainfall varying according to the daily records.	The individual time series for the eight primary gauges (introduced and analysed in Section 3.6.1.2) are used as inputs to the model. Thiessen polygons have been used to spatially distribute the precipitation data.

The Thiessen polygon method, as was also used in a MIKE SHE model by Vazquez et al., (2002) on the Grete catchment, Belgium, and Rubarenzya et al., (2007) is an acceptable choice of spatial representation due to the spatial variability in rainfall across the catchment being relatively small as a result of its predominantly low lying nature (as described in Section 3.2). The adopted approach is less labour and computationally expensive within the model compared to other more complex methods, yet still includes a method of spatially distributing the data.

4.2.4. Evapotranspiration

The reference evapotranspiration; the rate of evapotranspiration from a reference surface (grassland) with an unlimited supply of water, is used in MIKE SHE from which to further calculate actual evapotranspiration based on defined Leaf Area Index (LAI) and Root Depth (RD), described in the following Sections, 4.2.5 and 4.2.7, (DHI-WE, 2005). Within the Tern catchment models, reference evapotranspiration has been defined as spatially uniform, and uses the daily Hargreaves-Samani PET calculated in Section 3.6.2. Mc Micheal et al., (2006) also used Hargraeves-Samani derived Eto of their MIKE SHE model of semi-arid shrublands in California, and showed good results when compared to observed data.

Both the homogenous and distributed models use a uniform representation of Eto due to the sparse distribution of temperature data that was used in the Hargraeves-Samani calculation. It is acknowledged that uniform spatial representation of Eto is a limitation within the distributed model. However, the catchment average Eto (Section 3.6.2) that was derived for three locations (two of which are just outside the catchment boundary) showed very similar time series, therefore suggesting spatial variation in Eto across the catchment may be minimal and therefore is not a major factor in not being spatially variable in the distributed model.

4.2.5. Land use

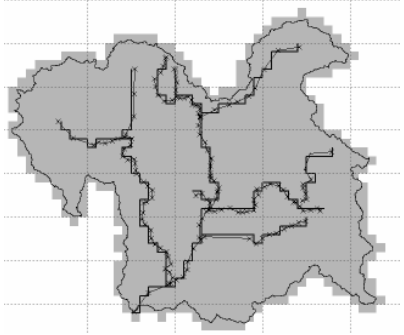
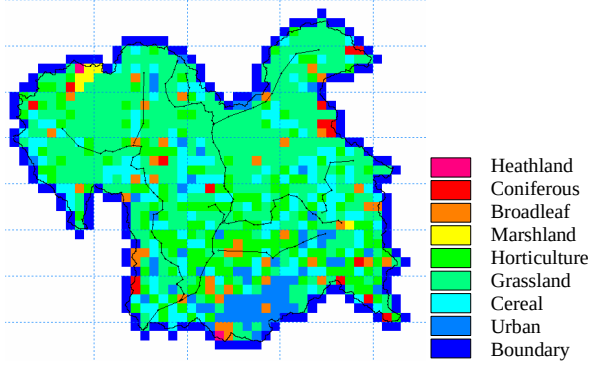
Land use is used by MIKE SHE to define the variations on the land surface that affect the hydrology within the model area, principally by specification of different vegetation types. For each class specified it is necessary to define associated Leaf Area Index (LAI) and Rooting Depth (RD) parameters that are used in the calculation of evapotranspiration, canopy interception and water volume available for use in the unsaturated zone and recharge to the saturated zone. The specification of land use also relates to the overland flow parameters such as the surface roughness coefficient.

Table 4.5 summarises the principle differences between the representation of land use between the homogenous and distributed models. As can be seen, the spatial representation of land use is specified as uniform in the homogenous model, with the associated LAI and RD values seeking to represent the catchment average pattern, by employing the values of the predominant land use in the catchment that is grassland (shown in Table 4.6).

In contrast, the distributed model represents land use with the eight main cover classes described and analysed in Section 3.5. Before incorporation to the model, the vector land use data were converted to gridded raster data and re-sampled to a 1 km resolution within ArcGIS. Where smaller land use parcels existed within the same 1 km cell, the dominant land cover became the represented type for the given cell. Although spatial complexity is lost within this process, the dominant land use areas are still represented, such as the urban class representing Telford, located in the south of the catchment.

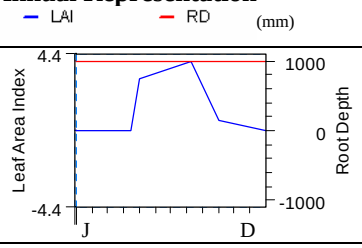
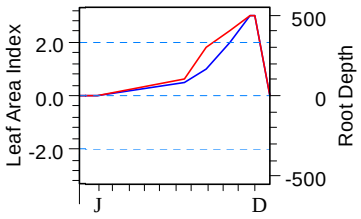
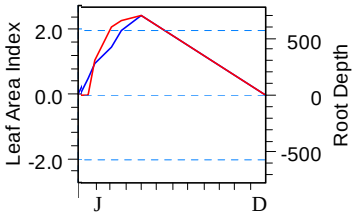
The associated LAI and RD values for each of the eight classes are shown in Table 4.6. Although many of the parameters are defined as constant (not varying with time), broadleaf, cereal and horticulture cover classes are associated with time-varying LAI and RD that fluctuate in accordance with an average growing season. The LAI and RD values in the Tern models are derived indirectly from the literature (e.g. Canadell et al., 1996; Schenk and Jackson, 2002), as commonly undertaken in other studies such as by Mc Michael et al., (2006) and Thompson et al., (2004). Some of the Rooting depths such as for potato and wheat, have been amended from the MIKE SHE default values for particular crop types DHI-WE, (2005), a method also adopted by Sahoo et al., (2006).

Table 4.5. Comparison of land use representation between the homogenous and distributed models

Tern Homogenous	**Tern Distributed**
	
The predominant catchment land use, grassland (40%), has been selected to represent the catchment as a uniform cover. MIKE SHE also requires Leaf Area Index (LAI) and Root Depth (RD) information for each cover class. LAI of 1.1 and Root depth of 100mm have been employed and correspond to the grassland values shown in Table 4.6 for the distributed model.	Land use is distributed in the model according to the eight classes defined in Section 3.5 and shown here in Table 4.6. For each land use, Table 4.6 summarises LAI and RD specified as time varying or constant, and gives the values defined for each class.

There is an inherent trade off between the two approaches used to represent land use in the models. In the catchment average setup of the homogenous model, spatial variation and heterogeneity are not considered, and as a result the representation of the canopy interception, evapotranspiration and water available for use in the unsaturated and saturated zones may be questionable. However, the aim of this model is in seeking to represent the average conditions of the catchment, which this conceptualisation does.

Table 4.6.Attribute parameters of land use in the homogenous and distributed Tern models (land use descriptions of agriculture type and woodland type were obtained from Ragg et al., 1984)

ID	Land use	Description	LAI	RD	Annual Representation
8	Broadleaf	Predominantly Oak, Ash or Birch	Varies 0 -4	Constant 1000mm	
7	Coniferous	Corsican Pine or Douglas Fir	Constant 2.65	Constant 1000mm	Assumed constant as pine needles are retained throughout the year
6	Cereal	Wheat or Barley	Varies 0-3	Varies 0-500mm	
5	Horticulture	Maincrop potato	Varies 0-2.5	Varies 0-700mm	
4	Grassland	For grazing, pasture	Constant 1.1	Constant 100mm	Assumed constant as grass cover is retained throughout the year
3	Heathland	Small up-stream areas of Roden, Tern and Meese	Constant 1	Constant 200mm	Assumed constant as the area represented by this class is relatively small
2	Marshland	Eg. Shropshire Meres (NW) and Aqualate Mere (E)	Constant 1	Constant 100mm	Assumed constant as the area represented by this class is relatively small
1	Urban	Areas of Telford, Wem, Shawbury, Market Drayton	Constant 0.5	Constant 100mm	Assumed constant and that urban areas are scattered with a small amount of vegetation

In contrast, Table 4.6 shows the distributed model includes LAI and RD parameters which vary with the different land use classes. Although this approach represents some spatial variation, a large degree of uncertainty is introduced to the model at this stage. The LAI and RD still remain average representations for each class (when in reality the

variation is likely to be significantly larger – ie. RD and LAI of grassland for example, are likely to vary across the catchment). In addition, there are no data for any change in the pattern of land use between years (ie. different crop types being grown from year to year, or details of any land left fallow) and so although acknowledged as uncertainties, there are no further improvements in the representation that can be included.

4.2.6. Overland flow representation

There are two options available within MIKE SHE to simulate overland flow; the finite difference approach or the semi-distributed overland flow zone approach. The finite difference method that uses diffusive wave approximation of the St Venant equations has been used in both Tern models, as the objective is to use the physically-based rather than conceptual representations where possible.

The Manning surface roughness coefficient, detention storage and initial water depth are required as input parameters in the model to calculate overland flow. The homogenous model uses uniform values for each of these parameters, as represented in Table 4.7 and Table 4.8, with parameter ranges thought to be representative for the catchment average. The distributed model is set to use the same detention storage value and initial water depth and are not spatially distributed. The Manning surface roughness is spatially distributed according to the land use pattern, with different land use codes assigned different surface roughness parameter values.

Table 4.7. Comparative representations of the surface roughness coefficient used in both models

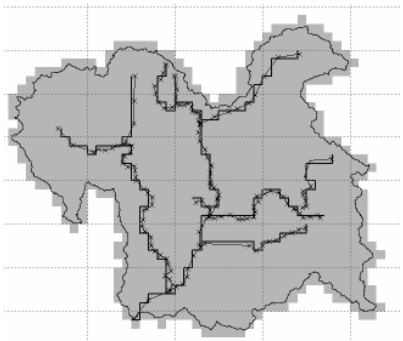
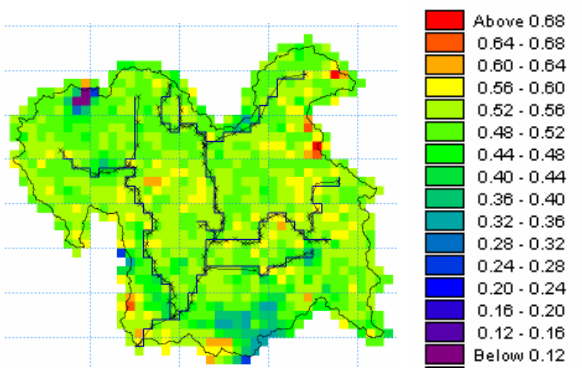
Tern Homogenous	**Tern Distributed**
	

Table 4.8. Comparative representation of overland flow in both models (** parameter values to later be defined by sensitivity analyses and model calibration)

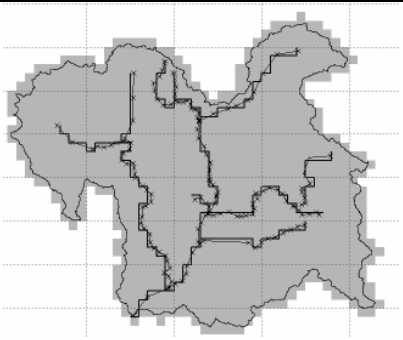
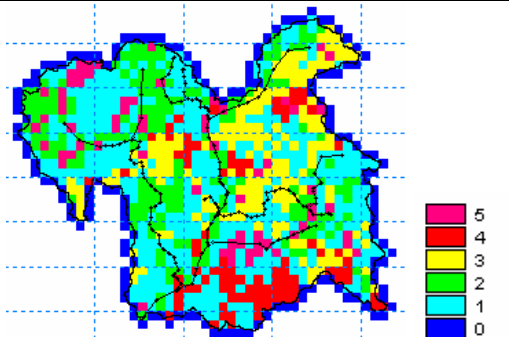
Parameter	Description	Representation and ranges in homogenous model	Representation and ranges in distributed model
Manning surface roughness	A flow resistance or roughness coefficient that determines the speed of overland flow. Equivalent to the Stickler roughness coefficient	Uniform** (typical range expected 0.1 to 1 based on values used by Vazquez et al 2008)	Defined for 8 classes based on land use types (section 4.2.5) (typical range expected 0.1 to 1 based on values used by Vazquez et al 2008)
Detention storage (mm)	To limit the volume of water allowed as overland flow. The value must be exceeded before OLF can occur	Uniform** (typical range expected between 0.005 – 0.3)	Uniform** (typical range expected between 0.005 – 0.3)
Initial water depth (mm)	The initial depth of water on the ground surface at the start of the simulation	Uniform** 0.005 used, though values over-written by hotstart initial conditions	Uniform** 0.005 used, though values over-written by hotstart initial conditions

4.2.7. Unsaturated zone

As reviewed in Section 2.4.2.4, there are a variety of methods that can be adopted in MIKE SHE to solve the unsaturated zone component of the model. The most commonly used methods are the physically based Richard's equation which is computationally demanding, and the two-layer water balance method that is less data, time and computationally expensive (DHI-WE, 2005). As an objective of this research is to compile uncertainty estimations from large numbers of model simulations and the HOST classification described in Section 3.4.2 to be of potential in reducing the number of soil classes within hydrological models (e.g. Dunn and Lilly, 2001; Marechal and Holman, 2005), the two layer approach has been selected to substantially reduce run-time.

The two-layer water balance method conceptually divides the unsaturated zone into a root zone and a below the root zone layer, with the objective of calculating a volume of water that is available at each time-step to recharge the saturated zone and also in the calculation of actual evapotranspiration. The approach draws on the input precipitation and evapotranspiration (that employs the LAI and RD as described above) and treats the hydrological response within the unsaturated zone as uniform in depth.

Table 4.9. Comparison of the different set-ups of the unsaturated zone in the two models

Tern Homogenous	**Tern Distributed**
	
Uniform spatial representation of the unsaturated zone parameters, shown in Table 4.10	Spatial distribution into 5 classes representing different HOST classes summarised in Table 4.11. The associated parameter ranges are shown in Table 4.10 and are obtained from secondary data from individual soil series

In addition to the precipitation, reference evapotranspiration, LAI and RD parameters, the two layer method requires the specification of a further four parameters, the soil water content at saturated conditions, field capacity, field wilting point and the infiltration rate (m s^{-1}), as summarised in Table 4.10.

Table 4.10 also indicates that in the distributed model five unsaturated zone classes have been specified. Details of these five classes are further detailed in Table 4.11 (which relates to the earlier presented Table 3.10). The five classes are defined based on a classification of HOST codes. For example, the unsaturated zone class 1 relates to HOST code 24, the dominant HOST code in the Tern catchment covering 42.4% of the area. Unsaturated zone class 2 comprises of slowly permeable soils with slight or seasonal waterlogging; defined by an amalgamation of HOST codes 18, 19 and 21. The HOST codes have been amalgamated in this class as they cover a smaller percentage of the catchment area and as a result reduce the total number of parameters required in the hydrological model. As a result of the five classes that have been defined, a total of 20 parameters are required to represent the unsaturated zone in the distributed model.

As the HOST system is a classification of soil types rather than a means of provision of hydraulic parameter values that are required by the model, reference has been made to the spatial extent of soil association data that were first shown in Figure 3.7 in order to derive the required parameter values for the unsaturated zone.

The required input parameter data have been extracted from the NSRI seismic report for the lead soil series of each soil association (Figure 3.7 and Table 3.8), for a land cover of permanent grassland.

Table 4.10. Comparison of the parameter values used for unsaturated zone representation **(values to be further assessed in sensitivity analysis and model calibration). All units are fractions unless specified.

Parameter	Description	Representation and ranges in the homogenous model	Representation and ranges in the distributed model
Soil water content at saturated conditions	The maximum water content of the soil (usually approximately equal to the porosity)	Uniform ** 0.345 – 0.592	Defined for 5 classes ** 1) 0.386 – 0.549 2) 0.345 – 0.526 3) 0.378 – 0.546 4) 0.354 – 0.592 5) 0.372 – 0.535
Soil water content at field capacity	The water content at which vertical flow is negligible (or the water content when the soil can freely drain)	Uniform ** 0.098 – 0.531	Defined for 5 classes ** 1) 0.290 – 0.456 2) 0.265 – 0.425 3) 0.145 – 0.370 4) 0.098 – 0.531 5) 0.105 – 0.394
Soil water content at field wilting point	The water content below which plants cannot extract water	Uniform ** 0.026 – 0.322	Defined for 5 classes ** 1) 0.140 – 0.309 2) 0.117 – 0.269 3) 0.034 – 0.154 4) 0.026 – 0.322 5) 0.028 – 0.171
Infiltration rate ($m\ s^{-1}$)	Saturated hydraulic conductivity of the soil	Uniform ** 4.05e-8 – 6.83e-8	Defined for 5 classes ** 1) 4.51e-8 – 6.37e-8 2) 4.05e-8 – 6.13e-8 3) 4.50e-8 – 6.37e-8 4) 4.05e-8 – 6.83e-8 5) 4.28e-8 – 6.25e-8
Depth of evapotranspiration (m)	The lowest elevation of the water table (the thickness of the capillary fringe)	Uniform ** 0.5 – 2.5	Uniform ** 0.5 – 2.5

These data report on the soil characteristics including the particle size fractions, organic and pH content as well as the required hydraulic properties of each lead series. The percentage volume of water retained at 0 kPA (Saturated conditions), 5 kPa (field capacity) and 1500 kPa (field wilting point) for the soil series within the Tern Catchment related to the parameters required within the model.

By overlaying the soil associations present in the catchment (Figure 3.7) with the 1km gridded HOST code map (Figure 3.8) in ArcGIS, it was possible to ascertain which soil associations were located within each HOST code. Reference to Boormann et al., (1995) also gave a percentage dominance of each soil association to each HOST code, with the dominant HOST code for each association used to represent the area.

The average values for each of the lead soil series (used to represent the soil association) within each HOST code were then calculated to give the parameter ranges in Table 4.10 for each class. For example, five soil associations are present within the HOST code 5 area in the Tern catchment (unsaturated zone class 5). The parameter values for class 5 are then derived by calculating the average saturated, field capacity, field wilting point and infiltration rate parameter values of the five lead soil series of each association. The result of this classification is a balanced method to derive parameter values based on representative soil characteristics, but within a hydrologically-related classification of soils to constrain the parameter values.

The differences in the two model setups are compared in Tables 4.9 and 4.10. The homogenous model is set to use uniform values throughout the catchment – with an attempt to best represent the unsaturated zone with one 'catchment average' parameter set. The parameter ranges defined in Table 4.10 compare the 'catchment average' parameter set (the upper and lower parameter limits defined as the minimum and maximum from the distributed setup just described).

Table 4.11. Summary of unsaturated zone classes and governing HOST classes from which they are defined

Class	HOST codes	Summary	Area (%)	Typical SPR (%)
1)	24	Slowly permeable soils with prolonged seasonal water-logging over slowly permeable substrates	42.4	40-47
2)	18, 19, 21	Slowly permeable soils with slight or seasonal water-logging over slowly permeable substrates	9.6	40-47
3)	3, 4	Relatively permeable and free draining soils on permeable substrate with deep groundwater	19.4	2-20
4)	9, 10, 11	Soils with shallow groundwater (within 1m depth) and artificial drainage	5.9	25-35
5)	5	Permeable soils that are free draining and on a permeable substrate with deep groundwater	21.7	2-12

Even though an attempt has been made to represent spatial variability in the unsaturated zone in the distributed model, due to the complexity of soils, and the variation in soil properties with depth and over short horizontal distances, the given representation is acknowledged as relatively crude as the unsaturated zone parameters are aggregated at a 1km grid scale, and for only five classes. In addition, uncertainty is introduced into the defined parameter values in the unsaturated zone because despite being from data for the specific lead soil series of each soil association, they are mean values from across the profile depth.

The representation of the unsaturated zone in the distributed model is however considered the best possible method given the data available and to avoid over-parameterisation. If the 32 individual soil associations had been specified as individual classes, there would have been a resulting 92 parameter values to specify for the unsaturated zone. As already noted, the method adopted with 5 unsaturated zone classes results in the specification of 20 parameters, with the definition that has been used based on a specific hydrological classification scheme (Boormann et al., 1995). Sahoo et al., (2006) also used a re-classification of the number of unsaturated zone classes within MIKE SHE, reducing the parameterisation from 31 to 16 classes.

4.2.8. Saturated zone

As presented in Section 2.4.2.5 there are two methods available within MIKE SHE to compute the saturated zone component of the hydrological models, the 3D finite difference method that uses physically based Darcian Law, and the conceptual linear reservoir method that uses interflow and baseflow reservoirs that gives an adequate yet simplified estimate of flow in the saturated zone. Both methods calculate the saturated sub-surface flow within a model, but with differing complexities.

As a result of the Shropshire Groundwater Scheme (Section 3.8.2), there has been substantial research undertaken by the British Geological Survey on the geology of the catchment, and as a result a large amount of data are available for use within this research. In addition, a principle objective of this research is to simulate groundwater levels (as well as river flow) and so for these reasons, and due to the complex structure of the geology of the catchment (as described in Section 3.3), the 3D finite method has been chosen as the most suitable method.

The 3D finite different method uses the 3D Darcy equation and is solved numerically by an iterative implicit finite difference technique (Graham and Butts, 2005). The method requires the specification of upper and lower levels of geological layers and lenses, the horizontal extents and further descriptive parameters for each lense including horizontal and vertical hydraulic conductivity, specific yield and storage coefficient, as shown in Table 4.12.

The homogenous model uses only one set of uniform parameter values for one geological layer set at a depth of -100m from the surface, with no additional lenses to describe the variation of the geology. The parameter ranges are defined as the minimum and maximum values from the limits defined for each class within the distributed model, as shown in Table 4.12. This homogenous representation of the saturated zone seeks to represent 'average conditions' within the catchment. It is acknowledged that this uniform specification perhaps limits the credibility of the model truly representing the physically based processes of the 3D finite difference method (Section 2.3). The method does not include the heterogeneity that is characteristic of the Tern catchment, that is comprised of substantial glacial drift deposits in the NW of the catchment and significantly ranging underlying solid geologies from low porosity Mercia Mudstones to porus aquifers' of the Sherwood Sandstones, as reviewed in Section 3.3.

In contrast, the distributed model adopts four classes of drift deposits specified as geological lenses that include glacio-fluvial sands and gravels and alluvial silty clay that inherently have very different hydraulic characteristics. The depths of these deposits are defined by drift thickness data provided by the British Geological Society (Figure 3.6). For use within MIKE SHE, the drift thickness data have been resampled to the same 1 km × 1 km grid as the other model files.

The drift lense thickness is defined in the distributed model as the top of the surface layer (topography in Section 4.2.2) minus the thickness of drift for each 1km cell in the catchment. A further solid geology lense layer comprising of six classes (Table 4.12) that include four sandstone and a low permeability and a mixed permeability class are then defined from the bottom level of the drift, to a bottom depth of -100m (where no drift is present the bottom layer of drift is zero and is therefore taken as the surface topography). As an example, if drift is present in a given cell within the catchment (e.g. ~40m depth as is possible in the NW of the catchment), then the represented solid geology lense would extend a further 60m beneath the drift layer to reach the bottom depth of -100m from the surface topography level.

Associated with the drift lenses and the solid lenses are the hydraulic parameters for each drift or rock type. The minimum and maximum allowed range of these parameters are shown in Tables 4.12. The ranges are derived from a combination of sources which

include the Tern geological infrastructure report (e.g. Allen et al., 1997 in LOCAR, 2000; Hobbs et al., 2002; Howard et al., 2008; Johnson, 1967).

Table 4.12. Conceptualisation of saturated zone parameters for the homogenous and distributed models and associated minimum and maximum parameter ranges
**(values to be further assessed in sensitivity analysis and model calibration)

Parameter	**Description**	**Parameter ranges for homogenous model**	**Parameter ranges for distributed model**
Horizontal hydraulic conductivity ($m\ s^{-1}$)	A water movement parameter defining the lateral rate of sub-surface flow	Uniform ** 1.0e-8 – 5.7e-4	Defined for 10 classes ** D1) 1.0e-5 – 9.0e-5 D2) 1.0e-5 – 9.0e-5 D3) 1.0e-4 – 5.7e-4 D4) 1.0e-8 – 1.0e-7 S1) 1.0e-8 – 5.0e-6 S2) 5.0e-5 – 9.0e-5 S3) 1.0e-5 – 9.72e-5 S4) 1.5e-5 – 9.72e-5 S5) 2.0e-5 – 1.736 e-4 S6) 1e-5 – 9.72e-5
Vertical hydraulic conductivity ($m\ s^{-1}$)	A water movement parameter defining the vertical rate of sub-surface flow	Uniform ** 1.0e-11 – 5.0e-5	Defined for 10 classes ** D1) 9.0e-9 – 5.0e-5 D2) 9.0e-9 – 5.0e-5 D3) 9.0e-9 – 5.0e-5 D4) 9.0e-9 – 5.0e-5 S1) 1.0e-11 – 5.0e-8 S2) 1.0e-9 – 5.0e-7 S3) 4.86e-8 – 1.04e-5 S4) 4.86 e-8 – 2.0e-5 S5) 5.68e-8 – 2.26e-5 S6) 5.78e-8 – 1.04e-5
Specific yield (fraction)	The drainable porosity indicating the volumetric fraction that a given aquifer will yield when all the water is allowed to drain out of it under the forces of gravity	Uniform ** 0.1 – 0.6	Defined for 10 classes ** D1) 0.1 – 0.3 D2) 0.1 – 0.6 D3) 0.1 – 0.3 D4) 0.1 - 0.3 S1) 0.1 - 0.35 S2) 0.1 - 0.35 S3) 0.1 - 0.35 S4) 0.1 - 0.35 S5) 0.1 - 0.35 S6) 0.1 - 0.35
Storage coefficient ($m\ s^{-1}$)	The capacity of the aquifer to release groundwater from storage in response to a decline in hydraulic head	Uniform ** 5.0e-5 – 0.3	Defined for 10 classes ** D1) 5.0e-5 – 0.25 D2) 5.0e -5 – 0.25 D3) 5.0e -5 – 0.25 D4) 5.0e -5 – 0.25 S1) 5.0e -5 – 0.3 S2) 5.0e -5 – 0.3 S3) 5.0e -5 – 0.3 S4) 5.0e -5 – 0.3 S5) 5.0e -5 – 0.3 S6) 5.0e -5 – 0.3

Where: D1= Till/diamicton, D2= peat, D3= glacio-fluvial sands and gravels, D4= alluvial silty clay, S1= low permeability mudstones, S2= mixed permeability class, S3= Bromsgrove/Helsby Formation, S4= Kidderminster/Chester Pebble Beds, S5= Bridgnorth/Kinnerton Formation, S6= Wildmoor/Wilmslow Formation.

Input saturated zone boundary data are not included as they have not been available, and are therefore set to zero for all cells on the boundary of the catchment. This may therefore be a cause of error in the water balance equation as discussed in Section 3.9. Groundwater movement from units that may extend outside of the catchment boundary (that is defined by surface water and topography) are therefore not represented.

Of all the hydrological components within the model set-ups described in this chapter, the saturated zone has the largest differences in spatial representation between the two models as the homogenous set up does not include lenses or different layers, and is uniform throughout the catchment to a depth of -100m. In contrast, the spatial representation of the saturated zone in the distributed model is technically quite complex. Nevertheless, the parameter values associated with the different lenses are unlikely to truly represent the processes because of the scale used; with a 1km × 1km grid cell size distributed by ten different geological lenses that have time constant values. However, the representation uses the data available as best possible so as not to outweigh complexity with realistic spatial representation in the model.

The parameter ranges shown in Tables 4.12 are a best representative set derived from the literature, and are considerably well researched based on the available data (e.g. LOCAR, 2000). However, the inclusion of this complex representation naturally results in further errors and uncertainty introduced in the models as it remains difficult to assess parameter values for some of the saturated zone classes. The mixed permeability solid geology class (2) and the storage coefficients for all lenses, for example, are a best-estimate and not based upon any field research carried out within the catchment or surrounding area. In addition, the arbitrary representation of bottom depth of -100m for the saturated zone may also add additional uncertainty and error to the models, as in reality the saturated zone and bedrock extend to a greater depth than has been represented in the models.

4.3. Channel flow (MIKE 11)

As discussed in Section 2.4.2.6, the MIKE 11 1D hydraulic model code has been coupled to the MIKE SHE framework (Graham and Butts, 2005; DHI-WE, 2005). This section describes the construction of the hydraulic model of the River Tern that represents the main channel and tributaries, and is the same model used within both of the Tern catchment models.

4.3.1. Simulation setup

A hydro-dynamic, unsteady state, River Tern MIKE 11 model has been created that runs on an hourly time step between 01/06/1998 and 31/12/2003. The simulation period is set to coincide with the MIKE SHE simulation, and although the MIKE 11 model can run independently of MIKE SHE, when it is coupled, the time step in MIKE SHE overrides the hourly time step and as a result the discharge and water level are only stored at a daily frequency, set at 09: 00.

4.3.2. Network delineation

The representation of the stream network within the MIKE 11 river model has been undertaken by digitising the primary surface water features, from OS Landline Plus data, available from Edina Digimap, as shown in Figure 4.1. A simplification of the network has been made, with only the primary channels included within the river model.

The shapefile of the refined stream network was then imported to the network editor in MIKE 11, and the network subsequently delineated by digitising points and connecting the branches together. During the branch connection, automatic chainages were allocated to each point in the network. Chainage is allocated from upstream (0m) to the downstream connection to another branch or the basin outlet.

a) b)

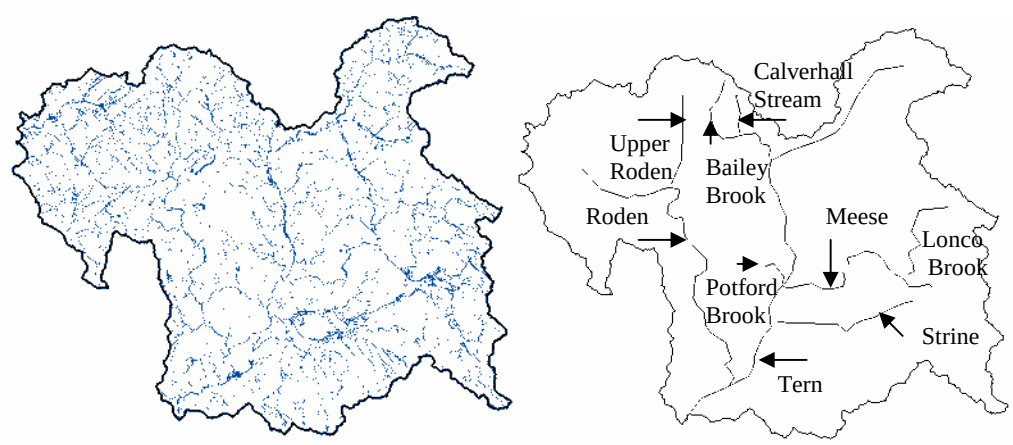


Figure 4.1. Comparison of a) all Ordnance Survey line water features and b) the refined stream network used in the river network delineation. Source a) © Crown Copyright/database right 2009. An Ordnance Survey/EDINA supplied service

On site visits to the catchment, it was noted that many of the other smaller water features shown in Figure 4.1a were negligible in size. For this reason, coupled with the issue of each of the river branches needing to be separated by the defined resolution of the hydrological model (1km x 1km cells) – the smaller streams were not included in the MIKE 11 river network. Additionally, a limitation in the licence used for MIKE 11 in this thesis was that only a set number of points were allowed to be digitised in the network, so it was also not possible to include many of the smaller streams.

The limitations of not including some of the smaller streams may be in the routing time being increased for water having to travel through the model to reach the larger streams and rivers. Also, there may be possible increases in the simulated groundwater levels throughout areas not near to the defined stream network in Figure 4.1b, since more water that might otherwise travel to the larger tributaries via smaller streams cannot do so as they are not represented in the MIKE 11 model.

It was noted in Table 3.25 and the conceptualisation of the catchment in Figure 3.44 that the River Strine in the south-east of the catchment is characterised as low lying with the smallest S1085 of all the tributaries of 13.6m, with artificial drainage of the Weald Moors. The predominant soil types in this part of the catchment (class 4, Table 4.11) have shallow groundwater with artificial drainage. Given this, the limitation of not including artificial drains in the MIKE 11 model is acknowledged and in this part of the Tern catchment likely to result in over simulation of groundwater levels (though as none

of the groundwater level boreholes are located in this region it cannot numerically be assessed).

Table 4.13 summarises the names, chainage and branch connections for each stream included in the model. The River Tern (42, 154m) and River Roden (31, 146m) are the longest rivers, and Potford Brook and Calverhall Stream defined as the shortest (3,008m and 3,398m) respectively. Table 4.13 also highlights that five of the eight defined branches drain into the River Tern, with the remaining being small secondary tributaries of the Roden, Bailey Brook and Meese. Notably, Potford Brook and the River Meese share their confluence with the River Tern at the same point of chainage of 26, 723m.

All of the branches are set to a kinematic routing method. This method is useful when the main concern is to route water to the main river system from upstream tributaries and other river branches (DHI-WE, 2005). The kinematic routing method is suggested as the best method of routing when using a MIKE 11 model within a MIKE SHE model, as it facilitates the use of large time steps that are often set by MIKE SHE (DHI-WE, 2005).

Table 4.13. Summary of branches, connections and chainages within MIKE 11

Branch Name	**Upstream – Downstream chainage (m)**	**Branch connection (m)**
Tern	0 – 42,154	-
Roden	0 – 31,146	Tern at chainage 36,832
Strine	0 – 12,382	Tern at chainage 30,261
Meese	0 – 16,871	Tern at chainage 26,723
Potford Brook	0 – 3,008	Tern at chainage 26,723
Bailey Brook	0 – 11,817	Tern at chainage 14,784
Lonco Brook	0 – 9,161	Meese at chainage 3,328
Calverhall Stream	0 – 3,398	Bailey Brook at chainage 7,598
Upper Roden	0 – 7,744	Roden at chainage 9,620

4.3.3. Cross sections

River cross sections of the River Tern main channel, its tributaries and floodplains have been extensively surveyed since the formation of the Shropshire Groundwater Scheme investigations (Section 3.9.2) and development of routine Environment Agency flood prediction models. These existing river cross-section data were available from the

Shrewsbury Environment Agency, West Midlands Region, in TIFF image format as original scans of the hand drawn surveyed sections.

The channels with cross section data are summarised in Table 4.14. It is highlighted that some of the cross section surveys are very old, dating back to 1990, and notably 1942-3 for Potford Brook. Despite the age of the data, and with the exception of Potford Brook, the data are considered more beneficial to use than to omit completely from the model. It is noted that significant changes may have occurred to the channels since surveys were undertaken, such as bank falls and silting up of beds, but these changes could likewise occur shortly after any of the more recent surveys.

The data were digitised from the scanned cross section TIFF images, and input to an excel database for every 1km interval on each channel. The majority of the field surveyed cross-sections were undertaken for the total floodplain width, up to 250m in length. However for the MIKE 11 set up, such large cross sections were not required and so cross-sections were digitised from left to right high flow banks.

Figure 4.2 shows the 41 selected cross sections (sampled every 1km upstream from the ~400 total cross sections available) on the main River Tern channel. The cross sections are plotted within the same figure to show how the nature of the channel varies with elevation. It is shown that the higher elevation cross sections nearer the source of the river are often narrower (~10m wide channels) and shallower (typically 2m from bank height to bed depth) than the lower elevation cross-sections located further downstream (typically 20-30m wide and 3 to 4m from bank height to bed depth). Figure 4.2 also shows three wider cross sections, these cross sections represent old water storage areas along the main stream. As reported in Section 3.7.1, sluice gates (such as those at the downstream storage area at Walcot) are no longer in operation and as a result have not been considered in the MIKE 11 model to actively store water or prevent its movement downstream. The inclusion of these larger cross sections does however allow the water to be held within a wider channel at these points.

Table 4.14. Summary of available cross-section data for channels within the Tern Catchment

Branch name	Survey dates	Surveyed by	Extent	Intervals between sections	Format of data	Further comments
Tern	1999 and 2003	Barry J. Lowe (1999), EA (Spring 1999), Total Surveys Ltd (June, 2003)	~400 sections from Outlet to Severn to upstream Market Drayton. Missing upstream.	Typically 100m (check)	Excel file & associated ESRI Shp file	Some storage area info also given.
Roden	Apr-Sept 1990	NRA for STR, USA	96 sections for main 33,500m. Upstream missing.	Typically 500m	TIFF Images of hand drawn surveys	
Strine	Dec 1999-Feb 2000	NRA for STR, USA, by Total Surveys Ltd	219 sections for 15,443m from confluence with Tern.		TIFF Images of hand drawn surveys	Currently missing downstream 7km (pages 1-26 of TIFFs)
Meese	Nov 1990	NRA for STR, USA, by Invar mapping surveys.	83 sections for 21,500m from confluence with Tern to Aqualate Mere	Irregular intervals but at least every 1km	TIFF Images of hand drawn surveys	
Potford Brook	1942-1943	A. Barton & Cy. E, for River Severn Catchment Board.	18 sections over 22,000 ft from confluence with Tern	Irregular spacing	TIFF Images of hand drawn surveys	No specified width (×) measurement. No coordinates. Units in ft. Very old. Not used.
Bailey Brook	Sept-Oct 1990	NRA for STR, USA.	Downstream 4,000m of ~13,000m channel, from confluence with Tern. Upstream missing	Typically 500m but some more frequent.	TIFF Images of hand drawn surveys	
Lonco Brook	1) Feb-Jun 1988 2) Oct 1990 3) Aug 1992	1) STW-USD 2) NRA for STR, USA. 3) NRA for STR, USA.	All three from confluence with Meese to 4,855m upstream at Whitelyford Bridge	1) and 3) mostly every 100m, 2) approx every 1,000m	TIFF Images of hand drawn surveys	3 independent surveys undertaken

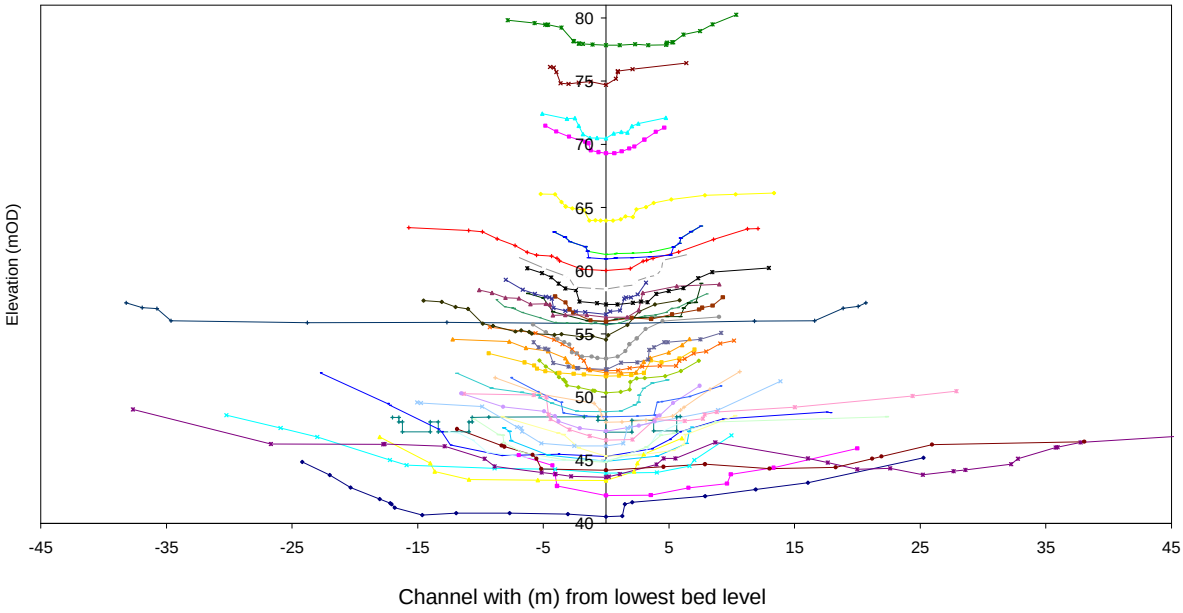


Figure 4.2. Comparative width and elevations of 41 selected River Tern cross-sections

In order to be able to specify cross-section data in the MIKE 11 model for the most upstream locations of the channels and tributaries, it has been necessary to synthesise cross-section data for areas in which surveys were not undertaken. Using a profiling methodology of Bailey Brook, an upstream tributary with existing cross section data, 11 cross-sections have been used to compute a representative mean cross section (Figure 4.3) that can be used as characteristic of other upstream sections of channels within the MIKE 11 model.

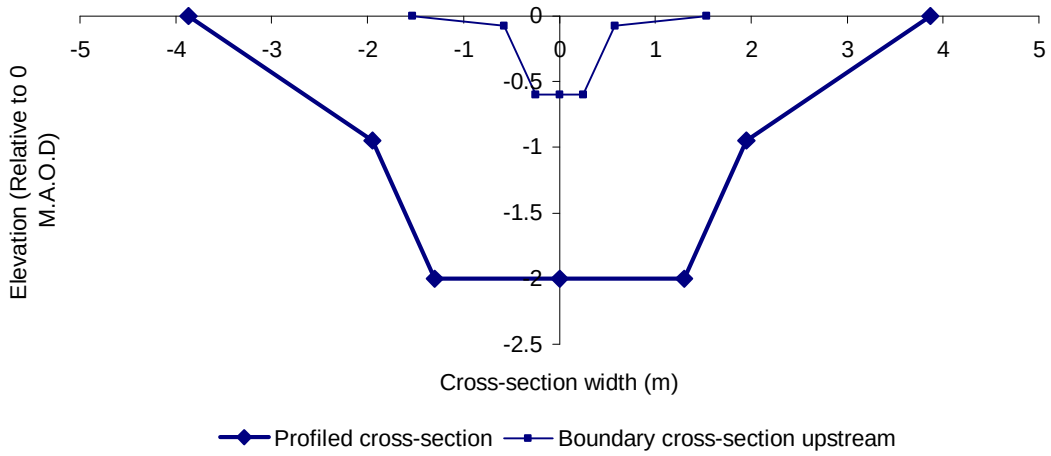


Figure 4.3. Comparison of synthetic profiled cross section and the reduced synthetic cross-section at the upstream boundaries

The synthetic cross-section has been further developed to account for changes in channel width and depth along the longitudinal profile. For every 1km upstream without

a surveyed cross section, a new synthetic cross-section has been created by reducing the channel width and depth data by 0.1m. The resulting upstream and smallest synthesised cross-section used for upstream beginnings of branches is also displayed in Figure 4.3. Figure 4.4 gives a summary of all the locations of the cross-sections input to the River Tern MIKE 11 model, in total 135 surveyed cross-sections, and 26 synthetic cross-sections.

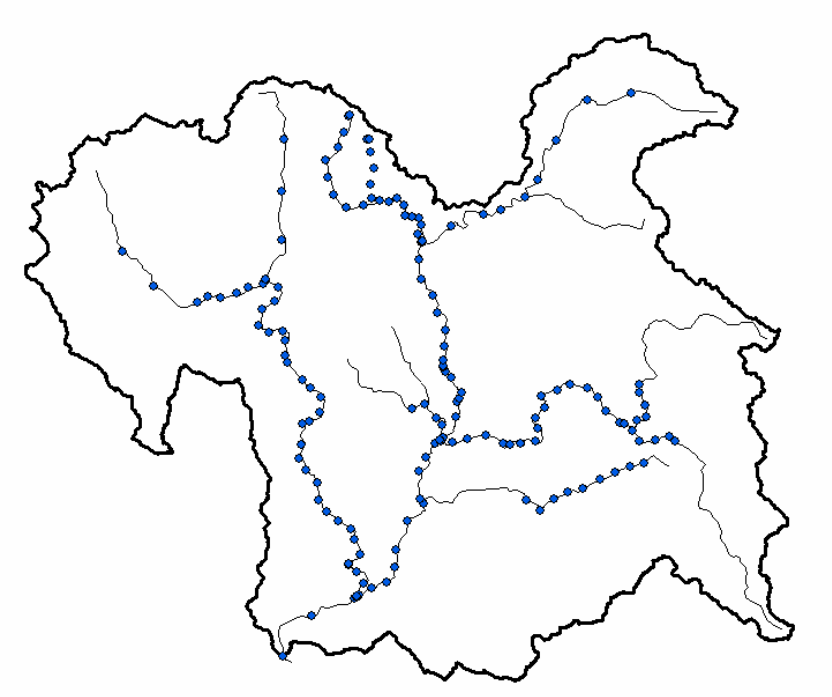


Figure 4.4. Locations of cross-sections input to the MIKE 11 river model

Slight modifications had to be made to the bed elevations at the downstream cross-sections of each branch, as they had to be manually adjusted as the same as the bed-values for the first cross-section of branch they connect to. This is a requirement of MIKE 11 to enable flow between channels as it is frequently the case that MIKE 11 has difficulty in interpolating bed heights between different branches.

Additional uniform hydro-dynamic parameters that are associated with the MIKE 11 model have been specified for the network. The Manning's coefficient for channel surface roughness is set to 30, a uniform value for every defined cross-section and thought to be representative for average catchment channels. Rather than adjusting the channel roughness coefficient for different branches (that would also increase the number of model parameters), the value is maintained as a constant as further information on the river beds (i.e. issues such as weed growth) were unknown.

Additionally, a bed leakage coefficient uses the MIKE 11 default value to allow a connection and water movement to occur between the channel and the surrounding sub-surface in the catchment. The completed river network and branches, cross section points (red), and upstream boundaries of each channel (blue) are shown in Figure 4.5.

In the individual profiles of each gauging station (Tables 3.22 – 3.29) influences on each gauge were discussed. It was noted that both licensed abstractions and effluent returns are likely to affect the stream flow of each tributary. Due to difficulties in acquiring any abstraction or return data, this is acknowledged as a further limitation in the MIKE 11 model representation of the Tern catchment. In discussing the extent of this limitation it is notable that the National River Flow Archive (NRFA) suggest (as shown in Tables 3.22 – 3.29), that the described abstractions and effluent returns appear to be in balance and not affect the overall volume of water in the catchment.

Despite this, it is unlikely to be the case that the abstractions and returns occur at the same time, with there likely to be strong seasonality involved in the summer abstraction of water for irrigation purposes. It may therefore be likely that in some of the more influenced tributaries (such as the Strine for example) that over simulation of river flow may occur during summer months when compared to the observed river flow data.

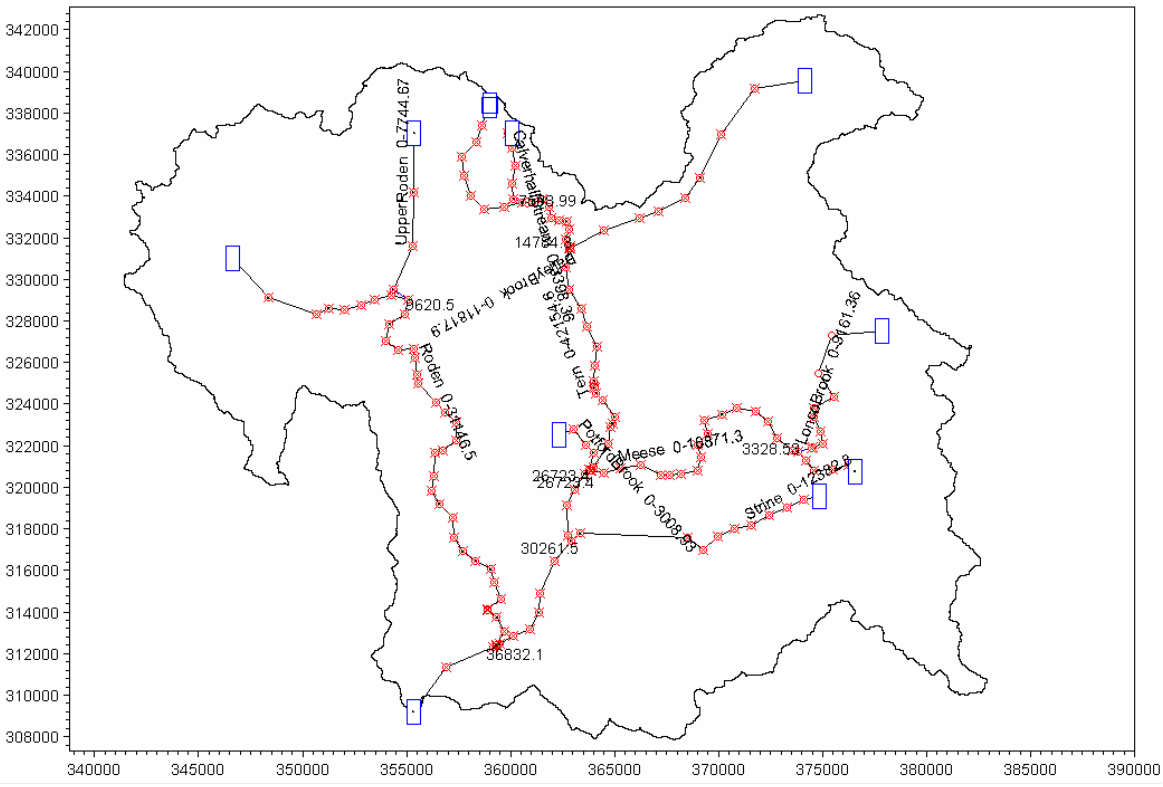


Figure 4.5. Network delineation in MIKE 11 Network editor

4.4. Summary

This chapter described the general set-ups of the two differing complexities of Tern catchment models in MIKE SHE – the Tern homogenous model and the Tern distributed model. The comparative methodologies of the setups of each hydrological component were simultaneously discussed to draw attention to the differences between the spatial representation of input data and parameterisation between the models. The construction of a MIKE 11 River Tern channel model that is used in both models was also described. The calibration (manual and automatic tests), validation, and rigorous sensitivity analyses of these fully operational models are subsequently presented in Chapter 5 for the homogenous model and Chapter 6 for the distributed model.

Chapter 5

Initial calibration, sensitivity analysis and automatic calibration of the homogenous Tern model

5.1. Introduction

Having detailed the set up of the two hydrological models in the previous chapter, this chapter focuses on the calibration, validation and a sensitivity analysis of the homogenous Tern model. Section 5.2 provides a description of the performance criteria by which the models are assessed including the development of a summary score measure used to quantify qualitative performance of the model. Initial simulations are then reported with the presentation of a supervised, manual calibration and validation of river flows and groundwater levels in Section 5.3. Section 5.4 details a sensitivity analysis of model parameters within the overland flow, unsaturated zone and saturated zone components of the model. The key parameters are identified that are then carried forward and subject to automatic calibration procedures in Section 5.5.

5.2. Performance criteria

In all model calibration and testing undertaken in both chapters 5 and 6, a specific range of performance criteria are used to assess the ability of the models. As introduced in Section 2.6.2, statistical measures are commonly used in catchment modelling that seek to quantify different aspects of a model's ability such as the overall performance using the NSE coefficient of efficiency (Nash-Sutcliffe, 1970) or the Root Mean Squared Error (RMSE), as well as more specific statistics that assess model performance in different parts of the hydrograph such as calculations of the error of peak flow or baseflow (Abu El-Nasr et al., 2005).

A research question detailed in Chapter 1 was to assess overall model performance not just with statistical measures but to also include more quantitative inspections of the hydrographs and groundwater levels, as well as more simplistic calculations such as the comparisons of observed and simulated mean daily flows, mean differences in groundwater levels and comparisons of total observed and simulated flows. Given this therefore, Table 5.1 summarises the performance criteria that are adopted as a means of model testing and the rationale and benefit for using each criteria.

With distributed hydrological modelling there can potentially be many stored output results for a given simulation. For example, output parameter results can be stored for every grid cell within the model, or river flow can be recorded at any node or point within the river network. These outputs can result in large volumes of data potentially available for inspection and assessment. For understandable reasons of vast repetition and time constraints, output results are not always carefully inspected, and instead, statistical summaries of performance are often preferred as a means of assessing model performance. As reviewed in Table 5.1, the statistical measures in general are good for fast comparison between different models and sites, but each have different weakness and bias that can mask what is shown in the real simulation when compared to the observed data. For example, the correlation coefficient, R, may be very good (>0.7) but the nature of the statistic only shows a relationship between two sets of data (e.g. observed and simulated model output), not how well the simulated data can replicate the characteristics of the observed data.

The task of ‘expert elicitation’ (Refsgaard et al., 2006) where subjective judgement and assessment of model performance and ‘degrees of belief’, are assigned to model simulations is acknowledged as a useful tool, given the knowledge that has been developed during the model conceptualisation (Chapter 3) and construction (Chapter 4) stages. Therefore, in addition to the independent statistically derived measures used in this thesis, an additional quantitative measure, ‘the summary score’, has been derived based on the amalgamation of some of the more qualitative and simple calculation statistics, as an example of expert elicitation. The rationale for including this measure is to include a quantitative score to the otherwise mostly qualitative information in the daily, monthly and regime plots of flows and groundwater levels.

Table 5.1. Performance criteria used to test the hydrological models of the Tern Catchment

Criteria	Measure	Category	Ability/what it shows	Caution/limitation
Plots of Daily flow	Flow at up to 8 gauging stations	Qualitative	Visual plots of observed v's simulated data from which to infer performance of baseflow, magnitude of peaks and timing of rise and falls in hydrograph.	Plots can appear noisy from daily variability making it harder to assess overall model performance.
Plots of monthly flow	Flow at up to 8 gauging stations	Qualitative	Simplifies the timeseries so that easier comparisons of observed and simulated flows can be made.	Masks the true variability seen in the daily record.
Flow regimes	Flow at up to 8 gauging stations	Qualitative	Compares the annual seasonal cycles giving a good overview of a model's all round ability.	Does not show if inter-annual variability can be reproduced by a model.
Plots of daily groundwater levels	Groundwater levels at up to 14 boreholes	Qualitative	Useful to infer overall performance and assess the daily pattern of groundwater fluctuations.	Observed data (in this thesis) are not daily, so direct comparison cannot be made.
Groundwater level regimes	Groundwater levels at up to 14 boreholes	Qualitative	Gives good overall assessment of whether the simulation can repeat the seasonal cycle.	Masks daily, monthly and inter-annual variability.
Rudimentary calculations	Flow and groundwater levels (8 gauging stations and 14 boreholes respectively)	Quantitative	Provides a good overview of the nature of the timeseries. Quick to calculate and easy to compare observed with simulated values.	No graphical representation of the timeseries to assess particular features (ie, timing).
Nash-Sutcliffe NSE	Flow at up to 8 gauging stations	Quantitative	Derived (initially) as a statistic for flow time series. Good general descriptor of overall ability and is dimensionless (can be compared for different sites and measures) .	As an independent statistic does not provide graphic detail of how a simulation visibly compares to the observed data.
Correlation R	Flow and groundwater levels (8 gauging stations and 14 boreholes respectively)	Quantitative	Indicates the strength and direction of a linear relationship between observed and simulated data.	Does not account for variance or consistent irregularities in the simulation when compared to observed. No graphic detail.
RMSE	Flow and groundwater levels (8 gauging stations and 14 boreholes respectively	Quantitative	Calculates the deviations of simulated points from their observed position, summing, then taking the square root of the sum. Gives the average magnitude of error, a good overall statistic to assess model ability. Commonly used.	Gives a high weight to large errors (i.e. peaks), therefore not treating the timeseries equally. No graphic detail.
RMSEP	Flow and groundwater levels (8 gauging stations and 14 boreholes respectively	Quantitative	Divides the RMSE by the observed mean daily flow or groundwater level to enable statistical comparison between sites (and make the statistic dimensionless).	As an independent statistic does not provide graphic detail of how a simulation visibly compares to the observed data.

The summary score is calculated separately for both flows and groundwater levels based on the different components shown in Table 5.2. As with the Correlation, R, and Nash-Sutcliffe, NSE, statistics, the summary score has a maximum highest value of 1 (that would pertain to a perfect model fit). The score for flows comprise of four components – with the baseflow ability, ability to reproduce peak flows and timing of events being subjective measures based on the poor, satisfactory and good rating scores shown. The ability to reproduce mean daily flow is a more objective measure based on the percentage error, but is still included within the scoring system to help reduce the subjectivity.

Table 5.2. Overview of the 'summary score' performance measure

Measure	Component	Description	Maximum score		Rating
River flow	Baseflow (qualitative)	The fit of the simulation during summer low flow months (JJA).	0.25	Overall score between 0-1 when adding each component score	Subjective score based on guide: Poor = 0.0 – 0.1 Satisfactory = 0.1 – 0.2 Good = 0.2 – 0.25
	Peaks (qualitative)	The fit of the magnitude of peak flows.	0.25		
	Timing (qualitative)	Whether simulated flows increase and fall on the same days as observed (irrespective of magnitude)	0.25		
	MDF (basic quantitative calculation)	How well the simulated MDF compares to the observed for the same time period	0.25		Objective score based on the guide comparing % error MDF: >20% error = 0.0 – 0.1 10 – 20% = 0.1 – 0.2 <10% error = 0.2 – 0.25
GW level	General level compared to observed (basic quantitative)	Ability of simulation to produce levels within the right magnitude	0.5	Overall score between 0-1 when adding each component score	Subjective score based on guide: >2.5m = Poor 1 – 2.5m = Satisfactory <1m = Good
	Overall shape compared to observed (qualitative)	Whether the variability (inter-annual and intra-annual) can be repeated by the model	0.5		Subjective score based on guide: Poor = 0.0 – 0.15 Satisfactory = 0.15– 0.3 Good = 0.3 – 0.5

The score for groundwater comprises of two components; the general ability of the simulated ground water level compared to the observed (measured based on a rating of general distance between observed and simulated data) and secondly, how well the shape of the simulated data corresponds to the observed data whether the seasonal fluctuations are reproduced in a similar way.

Each of the components in both the flow and groundwater measures are weighted equally, as in this research the aim is not to be able to produce a model that has better performance in simulating different aspects of the hydrograph (i.e. peak flows or baseflow), but to test the model's overall ability. In every case, the scoring has been undertaken blindly of other generated statistics, i.e. paying no attention to the NSE, R and RMSE in order to reduce bias. Additionally, when undertaking the scoring, each plot (for a given gauging station or groundwater level) were separated and judged independently so that performance at other sites would not bias the score for the one being assessed.

5.3. Initial manual calibration

This section describes an initial manual calibration of the homogenous Tern model that uses the set-up outlined in Chapter 4. The purpose of this initial calibration was to obtain an adequately calibrated model to then be carried forward for rigorous sensitivity analyses of the key model parameters (Section 5.4). The initial manual calibration was undertaken at a time when the operation and functionality of MIKE SHE were being learnt, and so rather than spending little time in this initial calibration then learning about model parameter response from the sensitivity analysis, considerable effort was given to learning how the variation of parameters affected the model outputs of flow and groundwater levels in a trial and error manner.

The completed set up and first simulation of the homogenous Tern model that was described in Chapter 4 initially used MIKE SHE default parameter values that are pre-set within the model, that were later re-defined for catchment-specific values. The purpose of this simulation was to ensure that the model could complete without any run-errors in the pre-processing stages. These first model simulations did not produce simulated hydrographs or groundwater levels that were calibrated to the Tern catchment.

The model parameters were then modified to more representative and specific values that fell within the defined parameter ranges that were provided in Chapter 4 (summarised in Table 5.3). The process of manual model calibration was iterative,

where parameters were perturbed gradually to consequently understand the response of model outputs. After each adjusted simulation, the flow at the basin outlet and other internal tributaries, as well as seven groundwater level plots were compared to the observed data, with the aim of trying to best fit the simulations to the observed data (and achieving best possible RMSE, NSE and R statistics), whilst maintaining parameter values within the limits defined. The calibration used the multi-proxy and multi-site method of assessment that was highlighted in Section 2.6.4.2 as necessary for all distributed model codes (Refsgaard, 1997; 2000).

Table 5.3. Parameter ranges, values and input data used in initial manual calibration of the homogenous model. ** denotes parameters assessed in the subsequent sensitivity analysis (Section 5.4)

MIKE SHE Component	**Parameter**	**Lower limit**	**Upper limit**	**Calibrated parameter value**
Model Domain	Grid size			1000m
Topography	Topography			1000m grid
Precipitation	Uniform according to daily Tern catchment areal mean			Time varying
Land use (uniform)	LAI			1
	Root depth			100mm
Evapotranspiration	Uniform according to calculated Hargreaves-Samani data			Time varying
Rivers and Lakes	MIKE 11			Constant
Overland flow	Manning number	0.1	1.0	** 0.50
	Detention storage			** 0.15 mm
	Initial water depth			** 0.005 mm
UZ flow	Soil water content at saturated conditions	0.345	0.592	** 0.50
	Soil water content at field capacity	0.098	0.531	** 0.40
	Soil water content at field wilting point	0.026	0.322	** 0.18
	Infiltration rate	4.05e-8	6.83e-8	** 6.12 e-8 ms^{-1}
	Depth of evapotranspiration	0.5	2.5	** 2 m
SZ flow	Lower level (relative to ground surface level)			-100m surface elevation
	Horizontal hydraulic conductivity	1.0e-8	5.7e-4	** 7.5 e-5 ms^{-1}
	Vertical hydraulic conductivity	1.0e-11	5.0e-5	** 2.0 e-6 ms^{-1}
	Specific yield	0.1	0.6	** 0.30
	Storage coefficient	5.0e-5	0.3	** 5.0 e-5
	Initial Potential head			-2m surface elevation

The selected parameter values in Table 5.3 were found to result in a visually and statistically robust model, not just at the basin outlet, but also for the other internal sites assessed. As shown, the defined parameter values all fall within allowed parameter limits. Notably the horizontal and vertical hydraulic conductivities are defined toward the higher part of the parameter range defined in Table 4.12. Knowing from the

catchment conceptualisation Figure 3.44 in Chapter 3 that sandstone is dominant in the basin and has a relatively high hydraulic conductivity, the calibration of the hydrologic conductivity to represent sandstone was considered an important characteristic to represent within the model. Sections 5.3.1 to 5.3.4 provide the results of this initial calibration and testing, for both river flows and groundwater levels, respectively.

5.3.1. River flow assessment

River flow has been calibrated with, and assessed at four key gauging stations within the catchment. These gauging stations represent four sub-basins within the catchment, as described in the catchment conceptualisation in Figure 3.44. Table 5.4 summarises the rationale for the selection of these specific gauging stations, with the predominant criteria being that they represent river flow from different tributaries and represent varied catchment sizes (as reviewed in Section 3.7) in order to assess intra-basin performance.

Table 5.4. Selection of river flow gauges used in model calibration

Gauge	Tributary/ Stream name	Description of stream	Catchment size (km^2)
Walcot	Tern	Encompassing the whole catchment for an overall view of model performance	852
Rodington	Roden	Draining predominantly mudstones of lower hydraulic conductivity	259
Tibberton	Meese	Draining predominantly sandstones of high porosity and hydraulic conductivity	168
Ternhill	Bailey Brook	A small relatively uniform upstream catchment of the Tern	34

Comparative plots of observed and simulated daily and monthly flows at the four gauging stations are presented in Figures 5.1 and 5.2 for the period January 2000 to December 2003. Inspection of the simulated daily flow hydrographs suggest a generally good simulation compared to the observed baseflows in all but Bailey Brook tributary, this is also confirmed by a comparison of observed and simulated average monthly river flows.

The highest peak river flows shown in the daily hydrographs at Walcot are not simulated as well as the baseflow, with winter flows during 2001-2 and 2002-3 being under-simulated. Conversely, in the very wet winter of 2000-2001 flow appears to be over-simulated, yet this appears averaged out when looking at the mean monthly flows at Walcot, as shown in Figure 5.2. This appears to give the impression that the hydrological model is good at representing the general tendency of wetter or drier years seen within inter-annual variability, however, the magnitude of peak flows are exaggerated. The pattern of under-simulation of peak flows (in all years) is also evident for the Roden and Bailey Brook tributaries. This is perhaps due to the horizontal hydraulic conductivity parameter being best representative of sandstone which is somewhat higher, 7.5 e-5 m s^{-1} (with minimal runoff therefore being generated) than should be expected from the less permeable mudstone of the Roden and Bailey Brook tributaries that perhaps have characteristically lower values of horizontal hydraulic conductivity (Howard et al., 2008; Hobbs et al., 2002). This also explains why the peak flow representation in the sandstone dominated Meese is much better.

The timing of the simulated hydrographs (with respect to rises and lags in the hydrograph limbs) are very good in the internal tributaries as well as at Walcot. This suggests that water is moving through the model at a comparable rate to that observed in the catchment.

Summary plots of observed and simulated annual hydrological regimes (2000-2003) for the four gauging stations are shown in Figure 5.3 for the hydrological year. The plots for the Tern at Walcot, Roden and Meese suggest a good reproduction of the regime, with the early spring (April) peak during the 2000-2003 period and the general low flow period between June-August simulated well. The annual flow regime is less well reproduced in December, with under-simulation at all gauging stations (likely a result of the already noted poor representation of peak flows during 2001-2 and 2002-3 winters). The simulated Bailey Brook annual flow regime re-confirms the reduced ability of the model in this small sub-catchment, although the correct regime pattern is simulated, the magnitude of flow is not well represented.

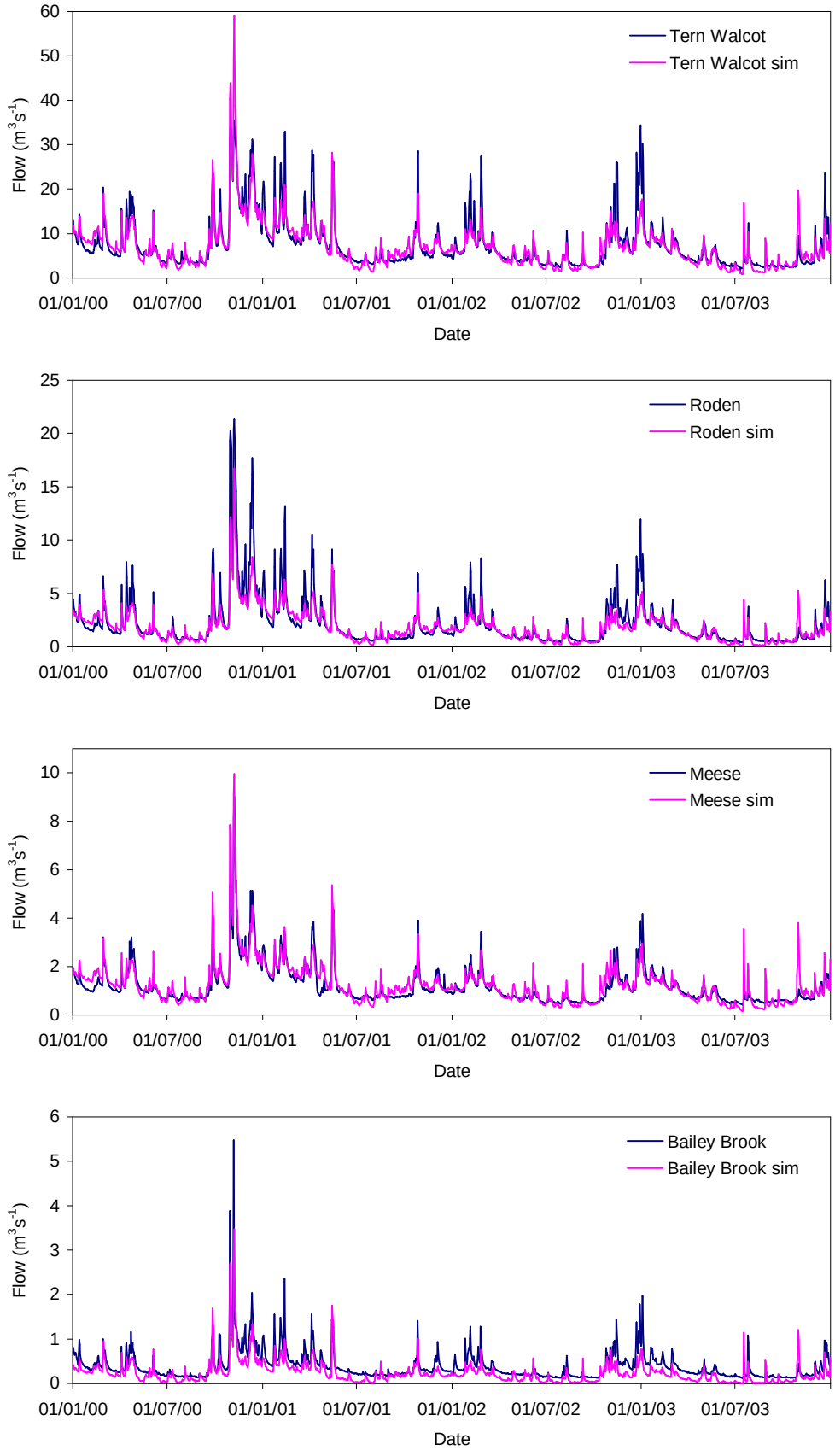


Figure 5.1. Observed and simulated daily river flows at four gauging stations used in the manual model calibration of the homogenous model

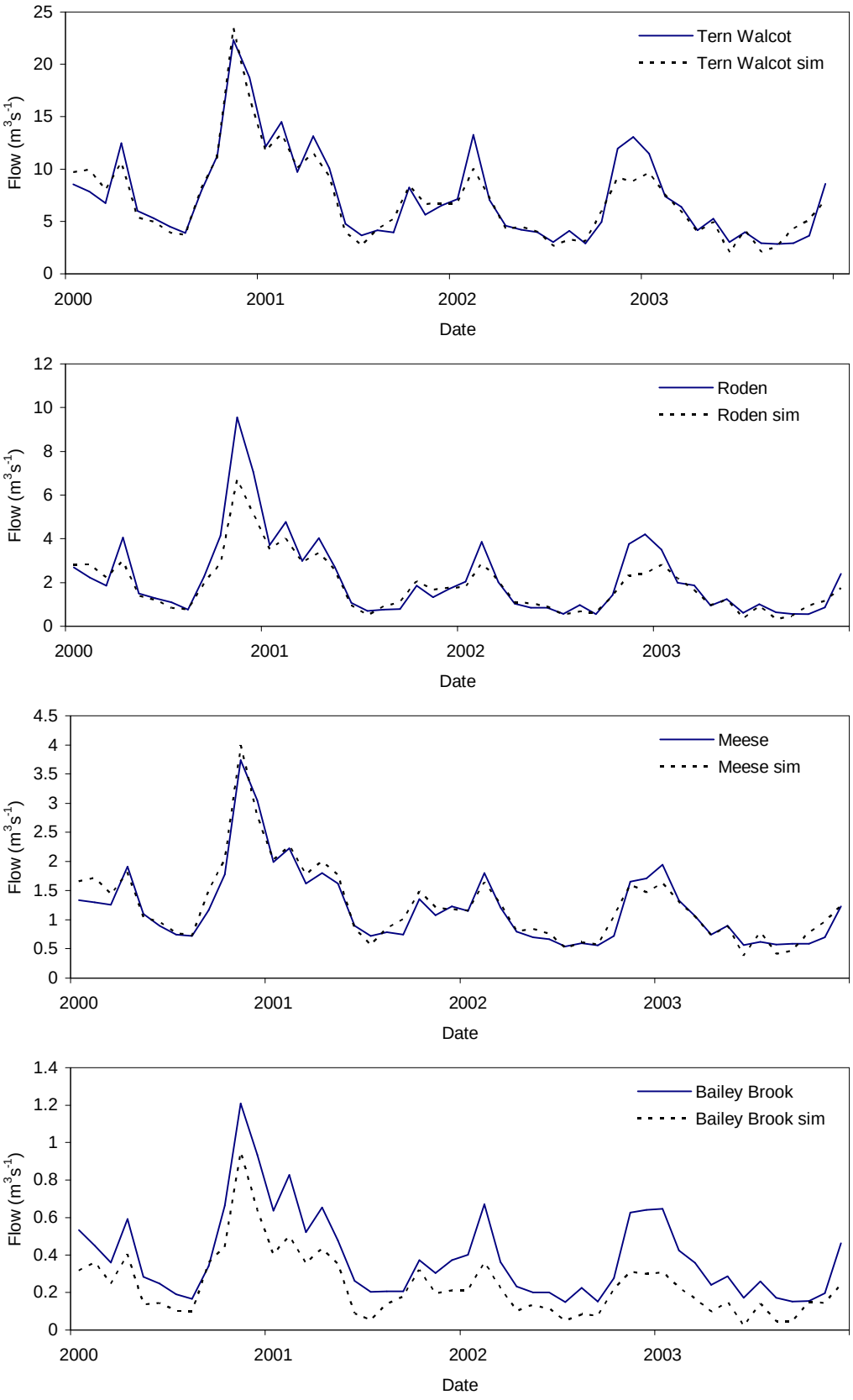


Figure 5.2. Observed and simulated mean monthly river flows for the four gauging stations used in the manual model calibration of the homogenous model

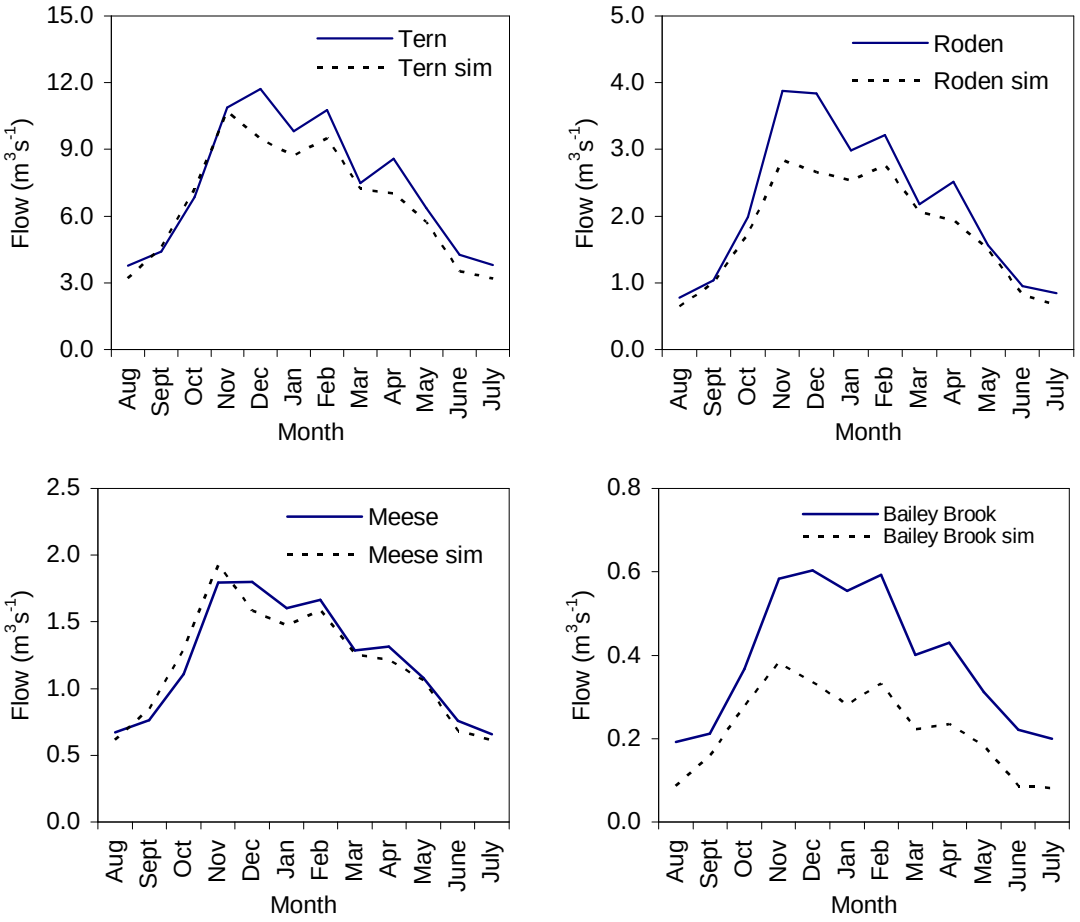


Figure 5.3. Observed and simulated annual river flow regimes (2000-2003) for the four gauging stations used in the manual model calibration of the homogenous model

The summary score statistic introduced in Section 5.2 has also been calculated for each gauging station. Table 5.5 shows the designated values for each component of the score, with the total for each gauge also given. The mean summary score is 0.695 for the whole catchment, compared to the 0.75 score at the basin outlet, a useful comparison to highlight that performance of the model when only examining at the basin outlet is different to that of including the internal performance of the model. The summary score system also rates the performance at the four sites in a similar way to the more quantitative statistics subsequently discussed and shown in Table 5.6.

Table 5.5. Results of summary score for manual river flow calibration of the homogenous model

Flow	**Walcot**	**Roden**	**Meese**	**Bailey Brook**
Baseflow	0.20	0.22	0.20	0.10
Peaks	0.10	0.15	0.16	0.15
Timing	0.25	0.22	0.22	0.20
MDF	0.20	0.12	0.24	0.05
Total	**0.75**	**0.71**	**0.82**	**0.50**
Mean score	**0.695**			

Further statistics for initial calibrated flow are given in Table 5.6. Mean daily flows are shown as consistently under-simulated in all four tributaries; this is likely due to an earlier recognised error in the water-balance addressed in Section 3.9 (with a likely over-estimation of evaporation, and an unknown movement of water into or out of the catchment within the sub-surface aquifers). It is also expected that the comparative total flows are subject to this input data uncertainty as every tributary is shown to simulate smaller volumes when compared to the observed.

The maximum observed and simulated river flows appear comparable for the internal tributaries of the Tern, yet at the basin outlet the maximum flow is somewhat over-simulated (Table 5.6), with almost twice the volume of water simulated at the peak event than was recorded. Although it cannot account for the total difference, it is expected that during this flood event, the Tern breached its banks at the Walcot site (as similar to the photograph in Figure 3.32), with the gauge thus unable to record the true peak flow, a problem that was also highlighted in the Walcot gauge review in Table 3.27.

Table 5.6. Comparative summary statistics of the manually calibrated homogenous Tern model (Observed values in bold, simulated values in normal text)

Gauge	**MDF** $\mathbf{m^3s^{-1}}$	**Max flow** $\mathbf{m^3s^{-1}}$	**Total flow volume** $\mathbf{m^3}$	**Correlation R**	**Nash-Sutcliffe NSE**	**RMSE** $\mathbf{m^3s^{-1}}$	**RMSEP**
Tern at	**7.37**	**35.44**	**9.31 x10^8**	0.81	0.74	2.87	0.39
Walcot	6.65	63.31	8.87 x10^8				
Roden at	**2.14**	**21.32**	**2.70 x10^8**	0.91	0.72	1.27	0.59
Rodington	1.75	17.98	2.34 x10^8				
Meese at	**1.21**	**9.02**	**1.52 x10^8**	0.88	0.75	0.41	0.33
Tibberton	1.18	11.84	1.58 x10^8				
Bailey Brook	**0.39**	**5.48**	**4.89 x10^7**	0.85	0.47	0.24	0.62
at Ternhill	0.22	4.11	2.97 x10^7				

The RMSE, correlation and Nash-Sutcliffe NSE suggest, overall, that the model performs very well at the basin outlet and for the two principle tributaries, Roden and Meese with Nash-Sutcliffe coefficients greater than 0.7, a value considered ‘good’ in a classification by Andersen et al., (2001) using MIKE SHE. It is notable that the highest correlation is shown on the Roden, 0.91, but this does not correspond to the highest NSE Nash-Sutcliffe coefficient which is seen on the Meese, 0.75, with also the best

RMSEP of 0.33. These differences may be due to the overall pattern and shape of the hydrograph being better for the Roden, yet the lower Nash-Sutcliffe likely a result of the magnitude of peaks being poorer at this site. The comparisons of the total flow volumes are useful to support this, with all sites but the Meese under-simulating total flow, where total flow volumes closely match the observed.

The overall poorer performance of Bailey Brook is reflected in the poorer statistics, with the RMSE almost as great as the mean daily flow and a RMSEP of 0.62 m^3s^{-1}. Again, this may be due to scaling issues, with the selected mean Tern catchment parameter values not being very well suited to the conditions found within this small mudstone dominated sub-catchment.

5.3.2. Groundwater level simulation and assessments

In addition to flow, seven groundwater level boreholes are also included in the initial manual calibration and assessment of the homogenous model. Table 5.7 summarises the rationale for the selection of these specific boreholes, with the predominant criteria being that they represent groundwater levels from different geologies (both solid and drift geology) and represent a good spatial distribution across the catchment area.

Table 5.7. Selection of groundwater level boreholes used in model calibration

Borehole name	Solid geology	Drift geology (and depth, m, if present)
Longdon	Bridgnorth/Kinnerton	Sands and gravels (10.38m)
Heathlanes	Bridgnorth/Kinnerton	None
Hawgreen	Bridgnorth/Kinnerton	None
Warren Farm	Kidderminster/Chester pebble beds	Sands and gravels (13.42m)
Edgmond	Kidderminster/Chester pebble beds	Diamicton (n/a)
Cherrington	Kidderminster/Chester pebble beds	Diamicton (n/a)
Gnosall *(Very close to catchment boundary)*	Wildmoor/Wilmslow	None

Comparative plots of observed and simulated groundwater levels are presented in Figure 5.4. All the y axes cover a range of 10m to enable the different sites to be compared on the same scale. The plots for these sites show differing levels of performance of the model to reproduce groundwater levels, although in general the pattern of water level fluctuation is good.

The model is shown to simulate groundwater levels at generally the right elevation (mOD) with Heathlanes and Warren Farm shown to always be simulated at higher levels, and Hawgreen mostly simulated at lower levels than typically observed. However, these level differences are mostly within ±2m of the observed levels. Notably, Cherrington and Longdon sites are simulated very well, with the general water level as well as the general seasonal cycle and shape of the simulation fitting well with the observed.

The plots are useful to assess as they highlight how well different seasons and specific recharge events during winter months are represented. In all sites, the peak recharge event with increasing groundwater levels during the wet 2000-2001 winter is simulated with generally the right timing and magnitude as the observed data. The exception to this is Gnosall, where the recharge event is not immediately apparent in the observed data (although it is noted that observed data are limited). To note more specifically, groundwater levels simulated at Longdon, Cherrington, Edgmond and Heathlanes sites are shown to have particularly comparable fits with fluctuations in the observed data. For example, the observed 2000-2001 winter groundwater level recharge of 1.66m at Edgmond and 1.41m at Cherrington are simulated with similar magnitudes of 1.85m and 1.87m, respectively.

The summary scores again reflect the discussion above, with the mean score rated as 0.66 for groundwater level simulation (Table 5.8), which although not as high as for river flow simulation still suggests a reasonable ability of groundwater representation in the model. Longdon and Cherrington are rated as highest scoring sites, both with 0.86, and Warren and Gnosall the lowest scoring with 0.50 and 0.48, respectively.

Table 5.8. 'Summary score' at the seven groundwater level boreholes used in the manual calibration of the homogenous model

	Cherrington	**Edgmond**	**Warren**	**Gnosall**	**Hawgreen**	**Heathlanes**	**Longdon**
General level	0.46	0.34	0.28	0.45	0.40	0.25	0.48
Overall shape	0.40	0.38	0.20	0.05	0.21	0.35	0.38
Total	**0.86**	**0.72**	**0.48**	**0.50**	**0.61**	**0.60**	**0.86**
Mean score	**0.66**						

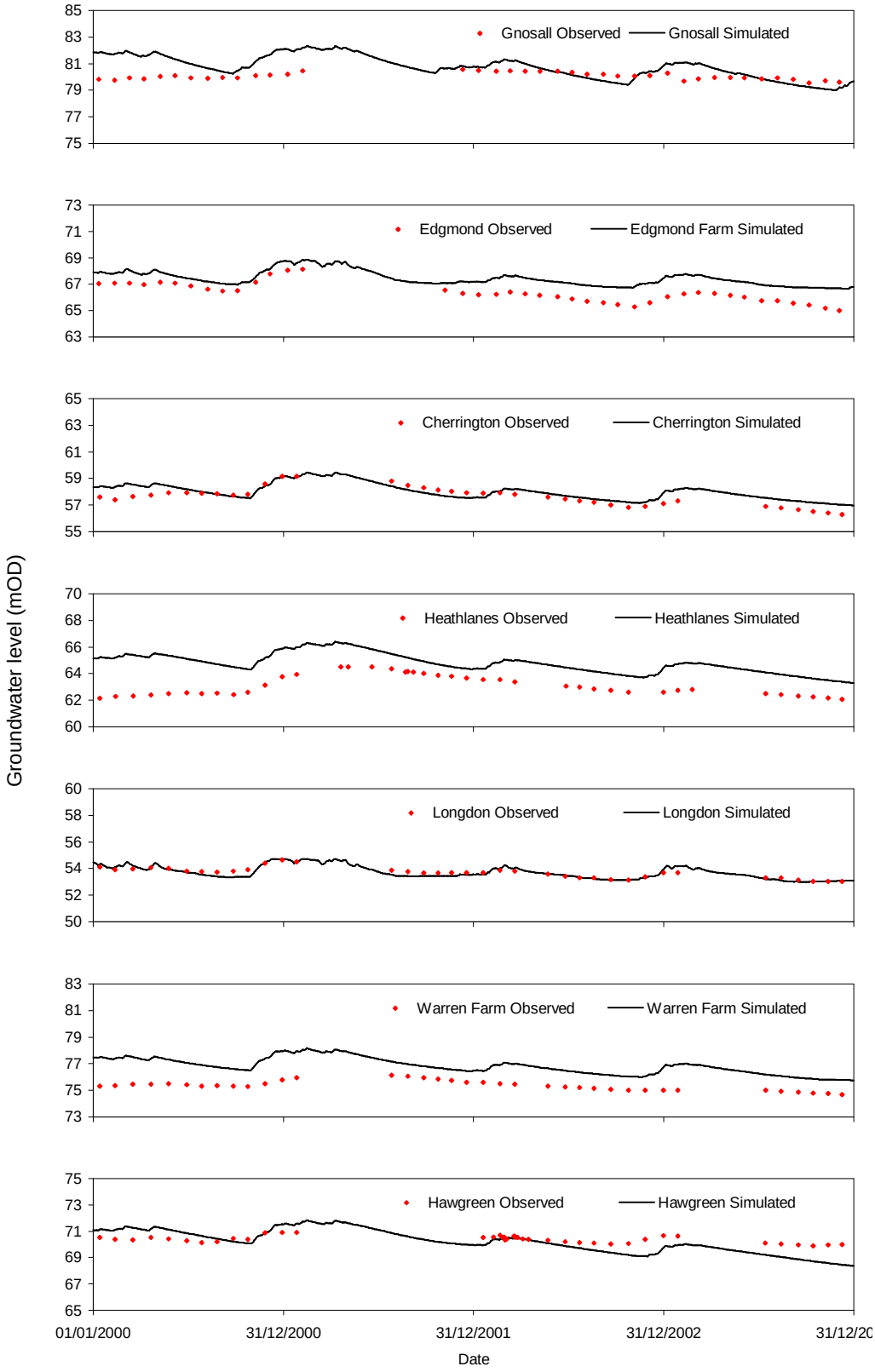


Figure 5.4. Observed and simulated groundwater levels at seven boreholes used in manual model calibration of the homogenous model

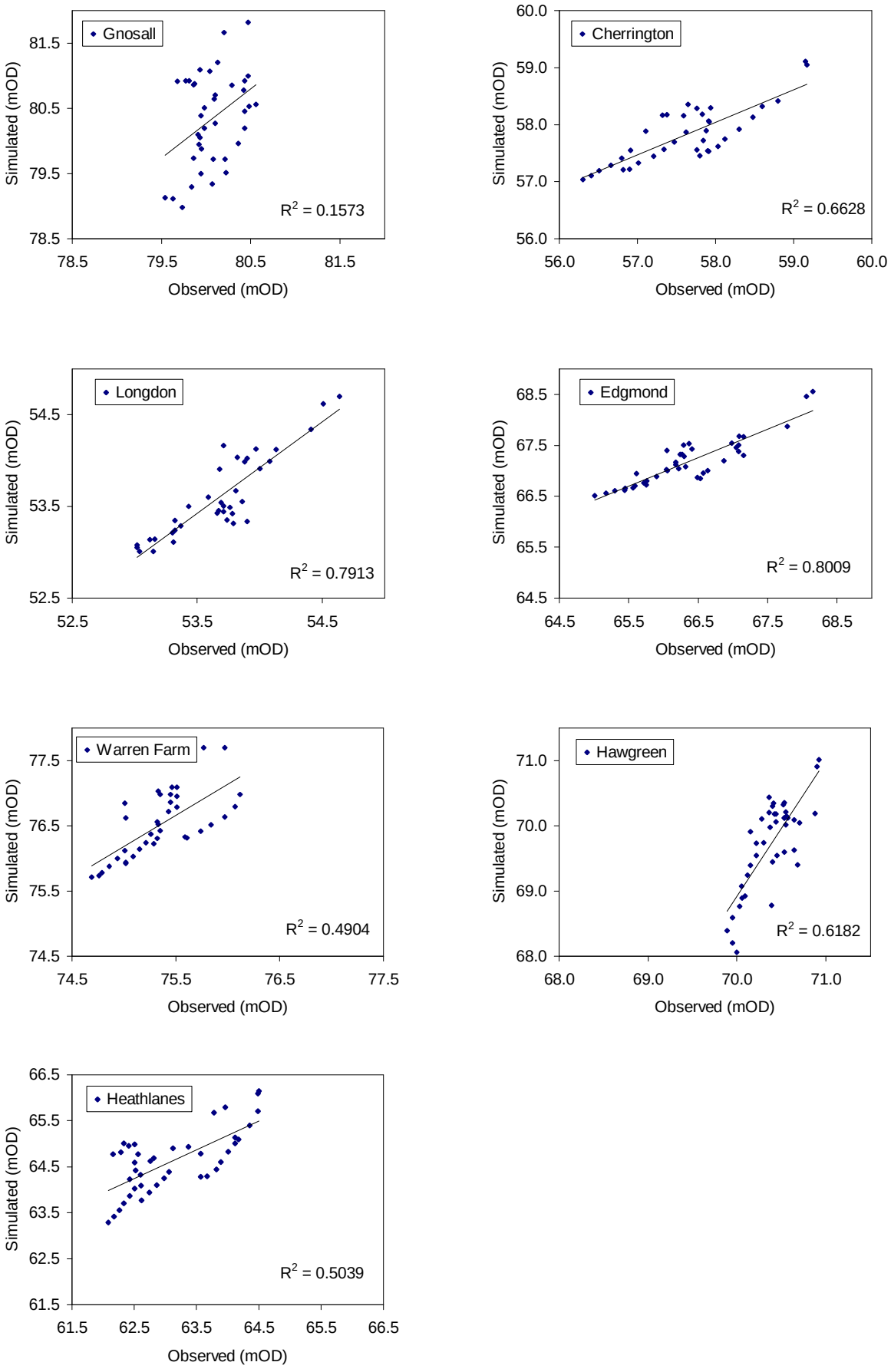


Figure 5.5. Relationships between observed and simulated groundwater levels for dates between 2000-2003 for the manually calibrated homogenous Tern model

Figure 5.5 provides a statistical and descriptive assessment of the relationship between the observed and simulated groundwater levels. The axes for each plot are set to the same ranges to enable an assessment of whether the simulation typically over or under simulates groundwater levels on the dates of observations. Ideally in a perfect model, the regression line should pass through x=y (1:1). The residuals from each regression line also indicate the spread and how well the simulated data fit the observed for each observation.

The R^2 at the sites varies between 0.16 at Gnosall, and 0.80 for Edgmond, thus indicating widely varying abilities of the model in simulating groundwater levels. Figure 3.39 indicated that Gnosall borehole is located at the very eastern extremity of the catchment and it is expected that this close to boundary location is possibly causing the difficulty in adequate simulation. The model is not set up to exchange any surface or sub-surface water along the external boundaries of the catchment, when in reality this exchange is likely to occur. It is therefore suggested that caution in ability to simulate groundwater levels at other close to boundary sites should be taken.

Although Edgmond has the best relationship between simulated and observed data, the plots indicate that the site at Longdon shows the trend line with the closest fit to x=y. This highlights a very good fit of the mean groundwater level of water being simulated close to the observed. Conversely, Hawgreen shows a satisfactory correlation, 0.62, yet consistent under-simulation of groundwater levels is seen with the range of simulated levels between 68 – 71m, yet the observed data levels falls no lower than ~70m.

Further statistics of the RMSE (m) and mean difference in observed to simulated levels (m) are shown in Table 5.9 for each borehole. These statistics help confirm the ability of the model of that seen in the plots, Figures 5.4 and 5.5, and the summary score in Table 5.8. Very good groundwater level simulation is seen at Longdon, with only a 7cm difference between mean observed and simulated data and RMSE of 0.22m. Heathlanes was previously flagged as consistently over simulating the groundwater level, as confirmed here with a mean difference of +1.52m.

Table 5.9. Statistical summaries of groundwater level simulations (for dates where simulations can be compared to observed between 01/01/2000 – 31/12/2003) for the manual calibration of the homogenous model

Site	Observed mean	Simulated mean	Difference (m)	RMSE (m)
Longdon	53.67	53.60	-0.07	0.22
Heathlanes	63.11	64.63	1.52	1.62
Hawgreen	70.39	69.73	-0.66	0.83
Warren	75.35	76.52	1.17	1.22
Edgmond	66.35	67.18	0.83	0.91
Cherrington	57.62	57.83	0.20	0.46
Gnosall	80.07	80.34	0.27	0.70

Considering the heterogeneous geology of the Tern catchment, the homogenous model appears to perform relatively well, given that the variability is represented with one set of parameter values to represent the 'catchment average'. It is noted that there are no observed data from the lower permeability Mercia Mudstone area to the north-west of the catchment, and that without these data it is difficult to fully comment on the ability of groundwater level simulation.

5.3.3. Validation of flows from the initial manual calibration

Overall, the manual calibration of the homogenous model has been shown to simulate both flow and groundwater levels with fair ability. This calibration is now tested in a validation process that uses additional un-used river flow and groundwater level observations at different locations within the catchment. Usually, the model should be tested with data from a different period using a split sample approach (Klemes, 1986) but with a short simulation period of four years it was considered more appropriate to use additional unseen multi-proxy data rather than split the calibration phase to any shorter time period (Refsgaard, 1997; 2000). In testing, the model outputs are presented and discussed for un-calibrated sites that used the same calibrated and un-adjusted parameter set shown in Table 5.3.

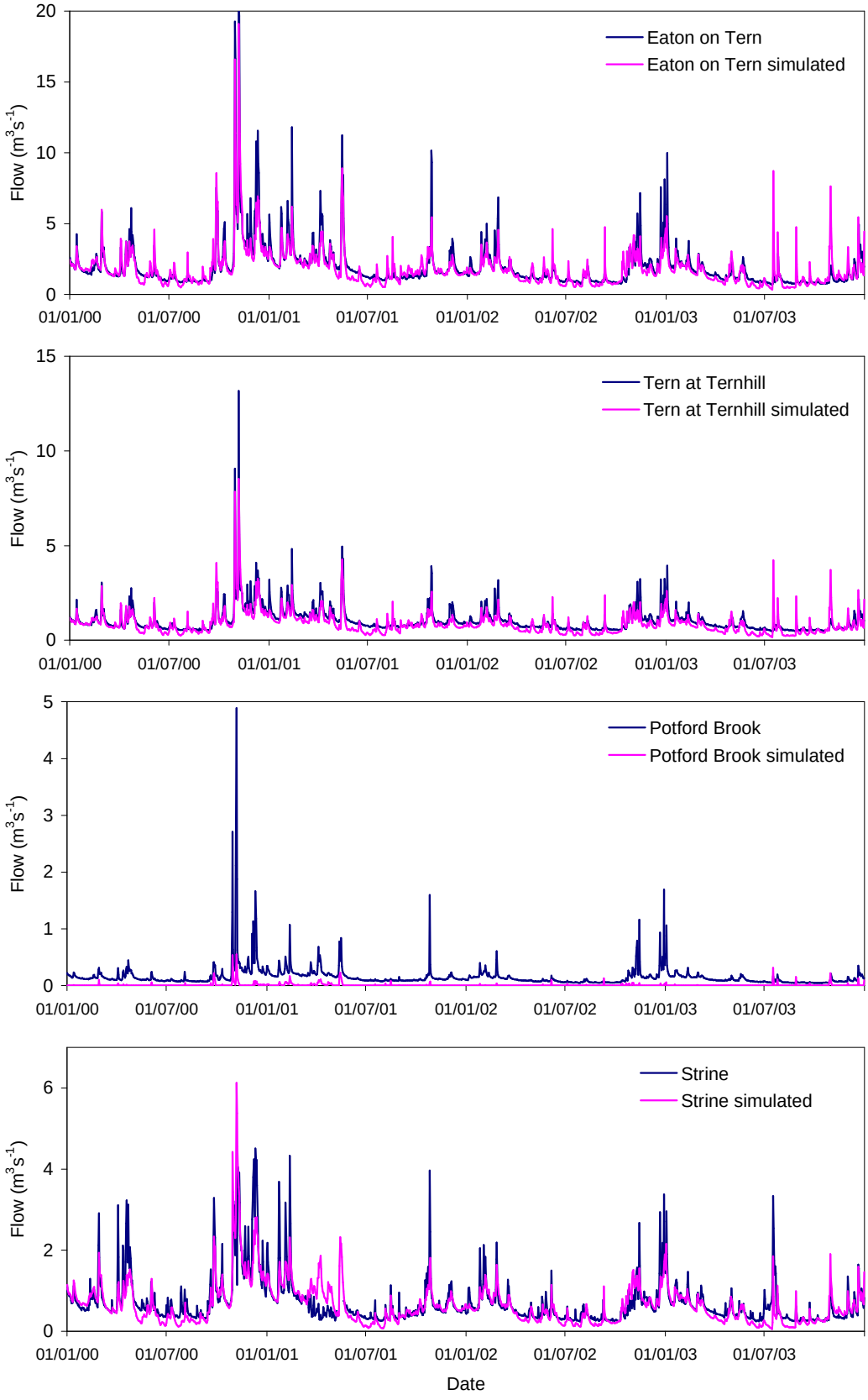


Figure 5.6. Observed and simulated daily river flows at four gauging stations in the model testing/validation of the homogenous model

The remaining four gauging stations, Eaton-on-Tern, Tern at Ternhill, Potford Brook tributary, and the Strine tributary were described in Section 3.7 and summarised in the catchment conceptualisation Figure 3.44. The first two of these sites are located on the Tern channel, upstream of Walcot. Potford Brook is the smallest represented tributary in the model with a catchment area of 25km^2, and the Strine is a low-lying channel with man-made drains (prone to more anthropogenic influence than the other tributaries).

Variable abilities of the model to simulate flow at the additional four gauging stations are shown in Figure 5.6, with the best model performance seen on the two additional Tern channel sites, Eaton-on-Tern and Tern at Ternhill. This is perhaps a result of the model already having been calibrated further downstream at Walcot on the same channel. At these two sites, baseflow and peak flows are simulated satisfactorily, although not as well as for the sites used in calibration.

For Potford Brook and the Strine, performance appears to be less satisfactory, and notably very poor for Potford Brook. Although the general peaks at Potford Brook occur at the correct time, the magnitude of the simulated flows are considerably under represented, and there appears to be no baseflow, with only flows simulated when there are peaks in the observed hydrograph. It is suggested that there is a scaling problem at Potford Brook that is causing the lack of simulated flow. As the catchment is quite small and represented by the fewest number of grid cells (with the MIKE 11 river model covering only three of them), perhaps less runoff is occurring between grid cells and as a result there is an overall smaller contribution of water entering the Potford Brook system. Inspection of the gridded result files suggests that due to the relatively coarse nature of the DEM, the direction of groundwater and overland flow are such that they enter the main Tern channel directly rather than being routed through the Potford Brook tributary.

Simulated flow in the Strine tributary appears to be in the correct magnitude as the observed data. What is notable however is that in May and June 2001 simulated peaks occur which are not evident in the observed data. As described in the set up of the MIKE 11 river model (Section 4.3), it is possible that water diversions or abstractions may have taken place during this time (that no data were available for and as such was described as a limitation of the model), as in the other observed data for the other tributaries peaks are also shown at this time. This is supported with the description of

the gauging station (Strine at Crudgington, Table 3.25) showing that of all the Tern tributaries, this is the most influenced by anthropogenic impacts.

Figure 5.7 displays the annual river flow regimes derived from the observed and simulated river flows, and Table 5.10 summarises the flow statistics. By assessing the monthly patterns in the annual regime 2000-2003 it is confirmed that in the Strine tributary, the summer months are under simulated compared to the observed data. Eaton-on-Tern and Tern and Ternhill show relatively good simulated flow regimes although once again all of the December flows are under-simulated as with the gauges used in the calibration process. Nonetheless, good Nash-Sutcliffe coefficients, 0.79 and 0.70 are calculated for Eaton and Ternhill respectively. Good correlation coefficients at all sites (0.70 – 0.90) support the plots of daily river flow in Figure 5.6 that appeared to show good simulation of the overall patterns of flow.

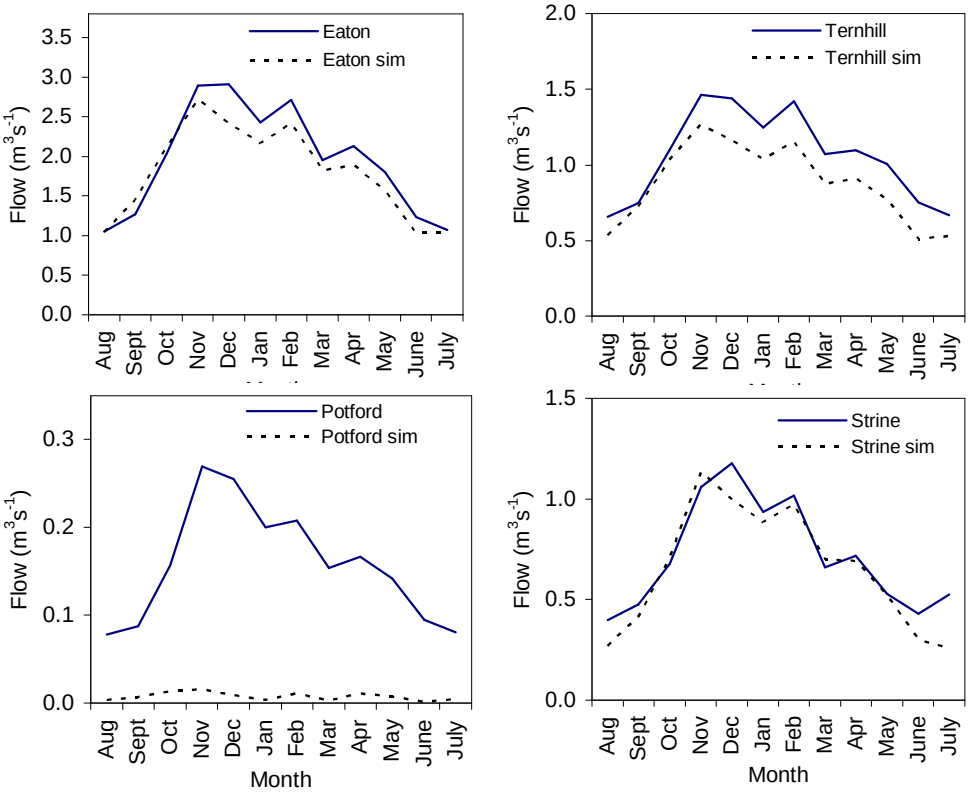


Figure 5.7. Observed and simulated annual river flow regimes at four gauging stations used in model testing of the manually calibrated homogenous model

The annual river flow regime for Potford Brook (Figure 5.7) is expectedly poor, as is the Nash-Sutcliffe NSE at -0.32, and RMSEP of 1.5, indicating there is no statistical ability of the model in reproducing flow at this site. In defence of the model, it is believed the daily flows in Figure 5.6 and regime in Figure 5.7 are useful tools, as they do show that the timing and variability of flow in Potford Brook are simulated with

some skill, it is the magnitude that is poorly represented. By only assessing statistical performance (using only the Nash-Sutcliffe, NSE, for example) this would have been overlooked.

Table 5.10. Summary statistics for the four river flow gauging stations used in testing of the manual calibration of the homogenous model. (observed values in bold, simulated values in normal text)

Gauge	**MDF m^3s^{-1}**	**Max flow m^3s^{-1}**	**Total**	**Correlation R**	**Nash-Sutcliffe NSE**	**RMSE m^3s^{-1}**	**RMSEP**
Eaton on Tern	**1.95**	**20.79**	**2.46×10^8**	0.90	0.79	0.74	0.38
	1.80	19.10	2.42×10^8				
Tern at Ternhill	**1.05**	**13.17**	**1.33×10^8**	0.88	0.70	0.37	0.35
	0.87	8.52	1.19×10^8				
Potford Brook	**0.16**	**4.89**	**1.99×10^7**	0.70	-0.32	0.24	1.5
	0.01	0.55	1.56×10^6				
Strine at Crudgington	**0.72**	**4.51**	**8.95×10^7**	0.79	0.60	0.37	0.51
	0.65	6.12	8.58×10^7				

Putting the poorly simulated river flow at Potford Brook into the context of the catchment performance downstream (in terms of total volumes of water), the poor simulation at Potford Brook is unlikely to result in a large difference to simulation at the basin outlet as observed MDF contribution (Table 5.10) is only 0.16 $m^3 s^{-1}$ of the 7.37 $m^3 s^{-1}$ observed at the outlet, and the total volume of observed water for the period is 1.99×10^7 which comprises only 2.14% of the total volume at Walcot.

5.3.4. Validation of groundwater levels from the manual calibration

In addition to the validation and testing of the calibration at additional gauging stations, seven additional groundwater level borehole records are available for the testing of the groundwater level simulation. The seven boreholes complement those in the calibration, again being from a range of geologies and being well spatially distributed across the catchment. As with the tested flows, the output plots and statistics are generated from the model that has been simulated with the same calibrated parameter set (Table 5.3), but previously un-assessed at these sites.

Figure 5.8 shows simulated groundwater levels for the seven sites along with the corresponding observed groundwater levels. As with flows, the model does not simulate additional groundwater levels that were not included in the calibration procedure to the

same ability. However, simulations at some sites are very good; notably at Woodlands Farm where both the general fluctuations and overall level are simulated well. Conversely, there appears to be no skill in the model at simulating groundwater level variability at Coley Farm or Cotton Farm, other than that the observed and simulated mean difference in groundwater level are within 1.77m and 2.13m respectively (as shown in the statistical summary Table 5.11). It is fair to suggest that these differences are within the same magnitude to many groundwater level modelling studies, especially in integrated studies where both flows and groundwater levels have been simulated (e.g. Refsgaard, 1997). It is also interesting to note that the Coley borehole is located close to the Gnosall borehole which was near the catchment boundary, where the model previously had difficulty in simulating observed levels. Additionally, the Coley borehole is known to be located within an area of groundwater abstraction (located in the more anthropogenically influenced south-east part of the catchment) which may be influencing the observed data.

The general fluctuation and groundwater level variability seen in the observed data are not simulated as well in this testing phase. With the exception of Woodlands Farm which as already noted has a good RMSE of 0.52m (Table 5.11), only the Heathcote borehole in 2002 and 2003 and the Radmoor borehole are able to adequately reproduce some of the annual variation. With regard to the large groundwater level recharge event in winter 2000-2001, the model simulates groundwater levels at Radmoor, Twinney and Woodlands Farm sites in accordance with the same magnitude as in the observed data, with an average 2.0 m recharge at most sites.

The overall performance RMSE measure suggests a range between the seven sites from 0.52m at Woodlands, to 2.13m at Coley Farm (Table 5.11). An explanation for the differing abilities at different sites may be due to the uniform parameter values used within this homogenous model. It is possible that the specified parameter set is more representative for some parts of the catchment than others, although within a catchment that characteristically has a complex geology of drift deposits in some parts (and of varying depths) as well as a wide ranging hydrogeology in the solid formation, it is difficult to assess.

To briefly summarise, this section has described the calibration and validation of river flows and groundwater levels at different locations within the Tern catchment. As could be expected, the model performs very well in some areas (especially within the calibration phase). The model also shows limitations, especially with simulating flows in the smaller tributaries which potentially may be as a result of scaling issues. The model shows inconsistency with the ability to simulate groundwater levels. At the Longdon, Cherrington, Heathlanes and Woodlands Farm sites, groundwater levels are simulated very well, yet at other sites, especially those close to the catchment boundary the models ability appears questionable. However, for a homogenous 'one parameter set' model applied to a mid-sized catchment, the simulations are considered good.

Table 5.11. Summary of groundwater level performance statistics for the seven boreholes used in the model testing phase

	Observed mean	Simulated mean	Difference (m)	RMSE (m)
Woodlands	70.21	69.86	-0.35	0.52
Radmoor	65.94	67.39	1.45	1.47
Heathcote	71.25	71.96	0.72	0.86
Rowton	61.61	61.54	-0.06	0.75
Cotton Farm	63.33	65.10	1.77	1.77
Twinney Lane	60.16	61.96	1.80	1.89
Coley Farm	71.04	73.17	2.13	2.13

5.4. Sensitivity analysis

This section details a manually undertaken sensitivity analysis that was performed for parameters in the homogenous catchment model to assess the ranges in simulated river flows and groundwater levels in different parts of the catchment, whilst varying individually the values of 12 parameters from three components of the model; overland flow, the unsaturated zone and the saturated zone. The purpose of the sensitivity analysis was threefold:

1) To undertake a model response test to the variation of parameters to assess the operation of the model, and check whether the perturbation of the parameters resulted in the expected, logical responses to simulated hydrographs and groundwater levels.

2) Given the initial manual calibration presented in Section, 5.3, to assess whether any of the sensitivity analyses results suggest the manual calibration could have been improved.

3) To quantify which parameter variations the model outputs were most sensitive to, and subsequently which parameters to carry forward for an automatic calibration procedure described in Section 5.5.

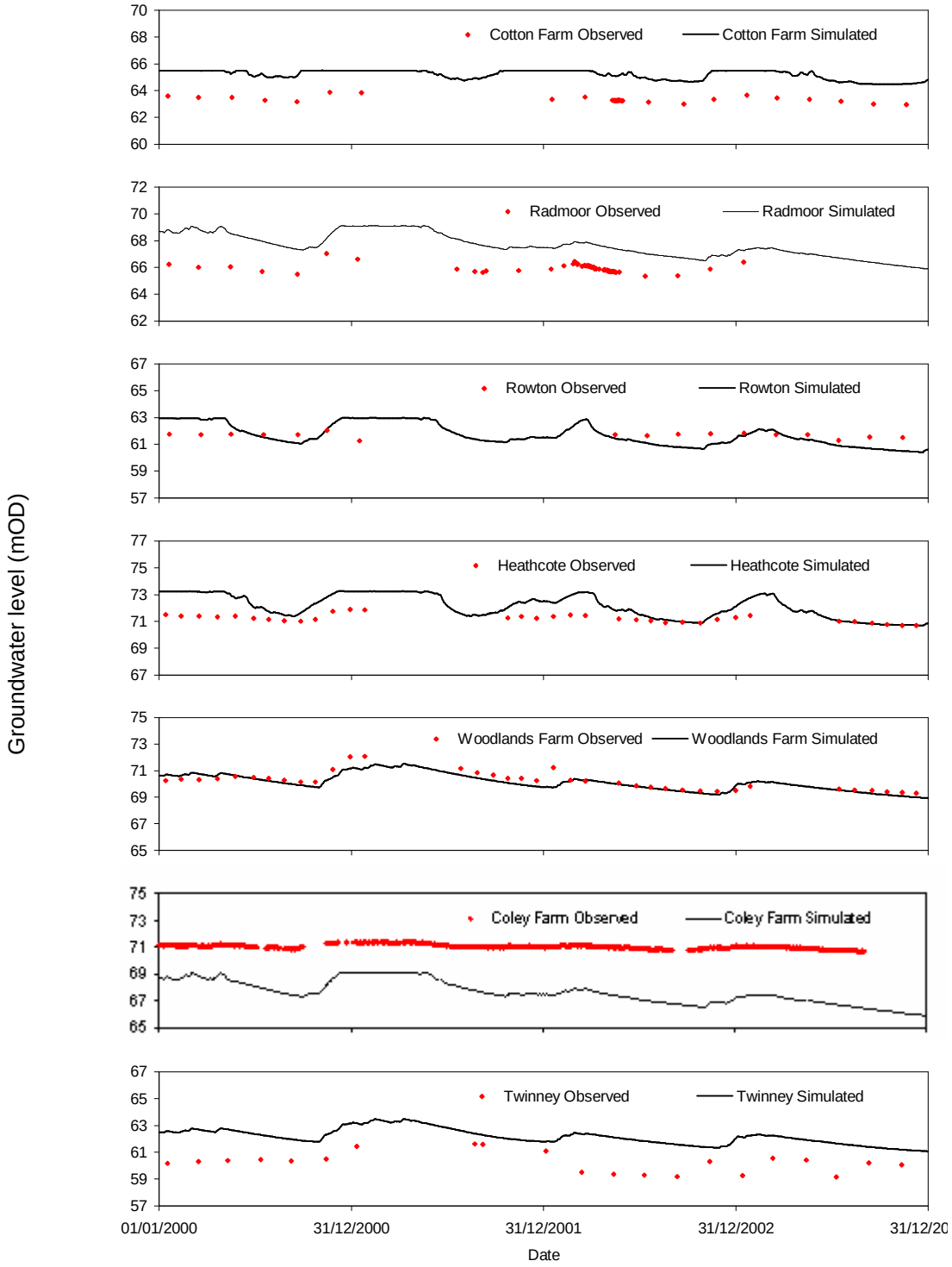


Figure 5.8. Observed and simulated groundwater levels at the seven boreholes used in the model testing for the manually calibrated homogenous model

5.4.1. Methodology and setup

As reviewed in Section 2.6.3.2, in order to calibrate a hydrological model most effectively, it is necessary to undertake a sensitivity analysis of the main model parameters. This is especially true with regard to automatic calibration procedures to reduce un-necessary over-parameterisation, computational time and complexity (e.g. Mertens et al., 2004).

Although unsupervised and automatic parameter sensitivity analyses can be undertaken (such as within the sensitivity analysis tool within the MIKE ZERO Autocal package) (DHI-WE, 2005), this sensitivity analysis adopted a supervised approach with the manual perturbation of selected parameters. This approach was chosen as it would provide a means of quantifying and displaying parameter sensitivity results that best compared to the manual modification method that was undertaken in the initial manual calibration (Section 5.3).

The selected parameters and their values assessed in the sensitivity analysis are summarised in Table 5.12 (with the highlighted control test '0' sharing the parameter values from the initial manual calibration described in Section 5.3). The initial manual calibrated value for each parameter are shown to have been perturbed by percentage factors at defined by intervals from a -99% change in the parameter value, to 99% change from the initial value. By changing each parameter value in both forward and backward directions from the starting value, the model sensitivity is assessed for both directions, a method suggested by DHI-WE, (2005).

As shown in Table 5.12, the percentage adjustment factors are at small intervals close to the initial calibrated value (e.g. 1% and 2% variations) whilst the intervals are greater at percentages further away from the initial value (e.g. 50% and 99%). This method allowed for more rigorous testing of the parameter values close to those already calibrated, to assess what effects small changes to parameter values close to the manually calibrated values would have (objective two of the sensitivity analysis). Although it would have been possible to perturb some of the parameters outside of the ranges selected (such as perturbation greater that 99%), the chosen method has been adopted in order to enable relative comparisons between parameters, so that the results can all be plotted on the same axes.

Each of the parameters were assessed individually; with each perturbation made, the model was then re-simulated and then outputs were extracted. The same parameter was then adjusted to the next percentage change, and the model re-run and outputs extracted, etc. In total, the initial Tern homogenous model was re-simulated 168 times, with 14 simulations for each of the 12 parameters assessed (though as Figure 5.12 indicates, SWC Sat and SWC Fc parameters could not complete simulations as the perturbed values were outside of feasible bounds).

The other input variables in the model (as shown in Table 5.3) were not perturbed or varied within the analysis, for example input precipitation, evapotranspiration and the land use data (as well as the other associated model parameters e.g. LAI and RD), and so are maintained as constants in each of the sensitivity simulations, to enable valid comparisons to be made for each parameter studied.

The results of parameter sensitivity on model outputs are assessed at four gauging stations at various key locations within the catchment (the same as used within initial model calibration, Section 5.3), not just at the basin outlet, Walcot. Detailed analysis of the time period of flows from October 2000 to September 2001 is selected as it includes a very wet winter season and a typically dry summer season. By including this range in conditions it is possible to assess sensitivity to both peak and low flows without including too much data in the daily plots where patterns might not otherwise be seen.

In addition to flows, groundwater levels at three sites (Gnosall, Cherrington and Longdon) were also assessed for the period 2000-2003. These sites were selected as they are located in three different geologies and from different parts of the catchment (Table 5.7, Section 3.8.1.2).

The following sections present the results and discussion of the sensitivity analysis. They include summary plots at each site for each component, for the mean percentage changes to flow and groundwater levels compared to the initial calibration, (control) ‘test 0’, as well as individual output results for parameters that the model is most sensitive to. Section 5.4.5 then reviews the sensitivity analysis process before automatic model calibration is undertaken in Section 5.5.

Table 5.12. MIKE SHE parameters selected for sensitivity analyses and ranges assessed.
(Values highlighted in **bold** were not included due to run errors for the specific value being used,
e.g. It being not possible for the model to run with values of SWC FC greater than that of SWC SAT)

Component	**Overland Flow**			**Unsaturated zone**					**Saturated zone**			
% perturbation of parameter value*	**Manning Surface roughness**	**Detention storage (mm)**	**Initial water depth (mm)**	**SWC Saturated (Sat)**	**SWC Field Capacity (Fc)**	**SWC Field Wilting Point (Fwp)**	**Infiltration rate ($m\ s^{-1}$)**	**ET Depth (m)**	**Horizontal HK ($m\ s^{-1}$)**	**Vertical HK ($m\ s^{-1}$)**	**Specific yield**	**Storage coefficient**
-99%	0.005	0.0015	0.00005	**0.005**	**0.004**	0.0018	6.12E-10	0.02	7.50E-07	2.00E-08	0.003	5.00E-07
-50%	0.25	0.075	0.0025	**0.25**	0.20	0.09	3.06E-08	1.0	3.75E-05	1.00E-06	0.15	2.50E-05
-20%	0.40	0.12	0.004	**0.40**	0.32	0.144	4.9E-08	1.6	6.00E-05	1.60E-06	0.24	4.00E-05
-10%	0.45	0.135	0.0045	0.45	0.36	0.162	5.51E-08	1.8	6.75E-05	1.80E-06	0.27	4.50E-05
-5%	0.475	0.1425	0.00475	0.475	0.38	0.171	5.81E-08	1.9	7.13E-05	1.90E-06	0.285	4.75E-05
-2%	0.49	0.147	0.0049	0.49	0.392	0.1764	6E-08	1.96	7.35E-05	1.96E-06	0.294	4.90E-05
-1%	0.495	0.1485	0.00495	0.495	0.396	0.1782	6.06E-08	1.98	7.43E-05	1.98E-06	0.297	4.95E-05
0	0.50	0.15	0.005	0.50	0.40	0.18	6.12E-08	2.0	7.50E-05	2.00E-06	0.30	5.00E-05
1%	0.505	0.1515	0.00505	0.505	0.404	0.1818	6.18E-08	2.02	7.58E-05	2.02E-06	0.303	5.05E-05
2%	0.51	0.153	0.0051	0.51	0.408	0.1836	6.24E-08	2.04	7.65E-05	2.04E-06	0.306	5.10E-05
5%	0.525	0.1575	0.00525	0.525	0.42	0.189	6.43E-08	2.1	7.88E-05	2.10E-06	0.315	5.25E-05
10%	0.55	0.165	0.0055	0.55	0.44	0.198	6.73E-08	2.2	8.25E-05	2.20E-06	0.33	5.50E-05
20%	0.60	0.18	0.006	0.60	0.48	0.216	7.34E-08	2.4	9.00E-05	2.40E-06	0.36	6.00E-05
50%	0.75	0.225	0.0075	0.75	**0.60**	0.27	9.18E-08	3.0	1.13E-04	3.00E-06	0.45	7.50E-05
99%	0.995	0.2985	0.00995	0.995	**0.796**	0.3582	1.22E-07	3.98	1.49E-04	3.98E-06	0.597	9.95E-05

*relative to run ‘0’ the initial manual calibration that is highlighted.

5.4.2. Overland flow

Summary plots of percentage change in mean river flow or mean groundwater levels from the initial calibration (control) are shown for each gauging station/borehole in response to variations of the overland flow parameters in Figure 5.9. Where the percentage change varies significantly on the y axis then the model can be described as highly sensitive to the parameter variation (named 'sensitivity test ID' on the x axis). Conversely, if the percentage range is minimal on the y axis, then the model is less sensitive to the parameter perturbation and hence is not effective as a parameter in model calibration.

Of the three overland flow parameters assessed, only the Manning roughness coefficient M (the inverse of the commonly used Mannings n) shows notable sensitivity upon model outputs, and only at the highest percentage variations to the parameter value (+/- 50% and 99%). Both the detention storage and the initial water depth variations show negligible effects on output river flow and groundwater levels. The initial water depth was included as a test parameter to ensure that negligible effects on outputs were given, as this parameter only affects the initial results at the start of the simulation (and in the case of this model setup, this is over-written with the iterative hotstart parameter values). The expected result of the model is therefore confirmed in this case.

Further detailed plots of simulated river flow and groundwater levels are included for variations to the surface roughness parameter in Figure 5.10 (in the interests of brevity the plots for detention storage and initial water depth are not included as they show little or no variation). The model shows a similar response to the variation of the Manning coefficient at all sites, with the -99% change of the roughness coefficient (a value of 0.005 as shown in Table 5.12) to result in extremely smoothed hydrographs without any peaks.

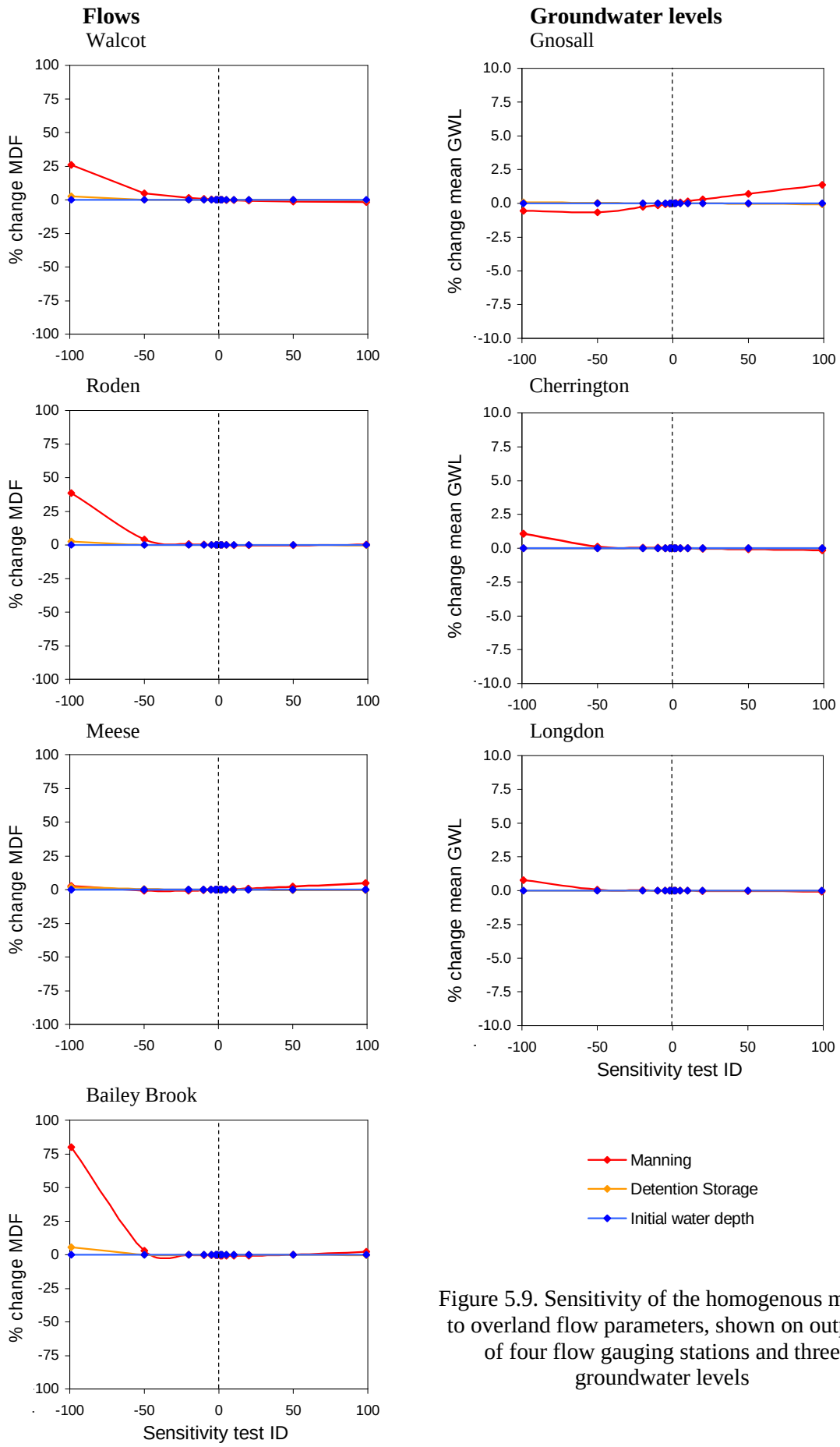


Figure 5.9. Sensitivity of the homogenous model to overland flow parameters, shown on outputs of four flow gauging stations and three groundwater levels

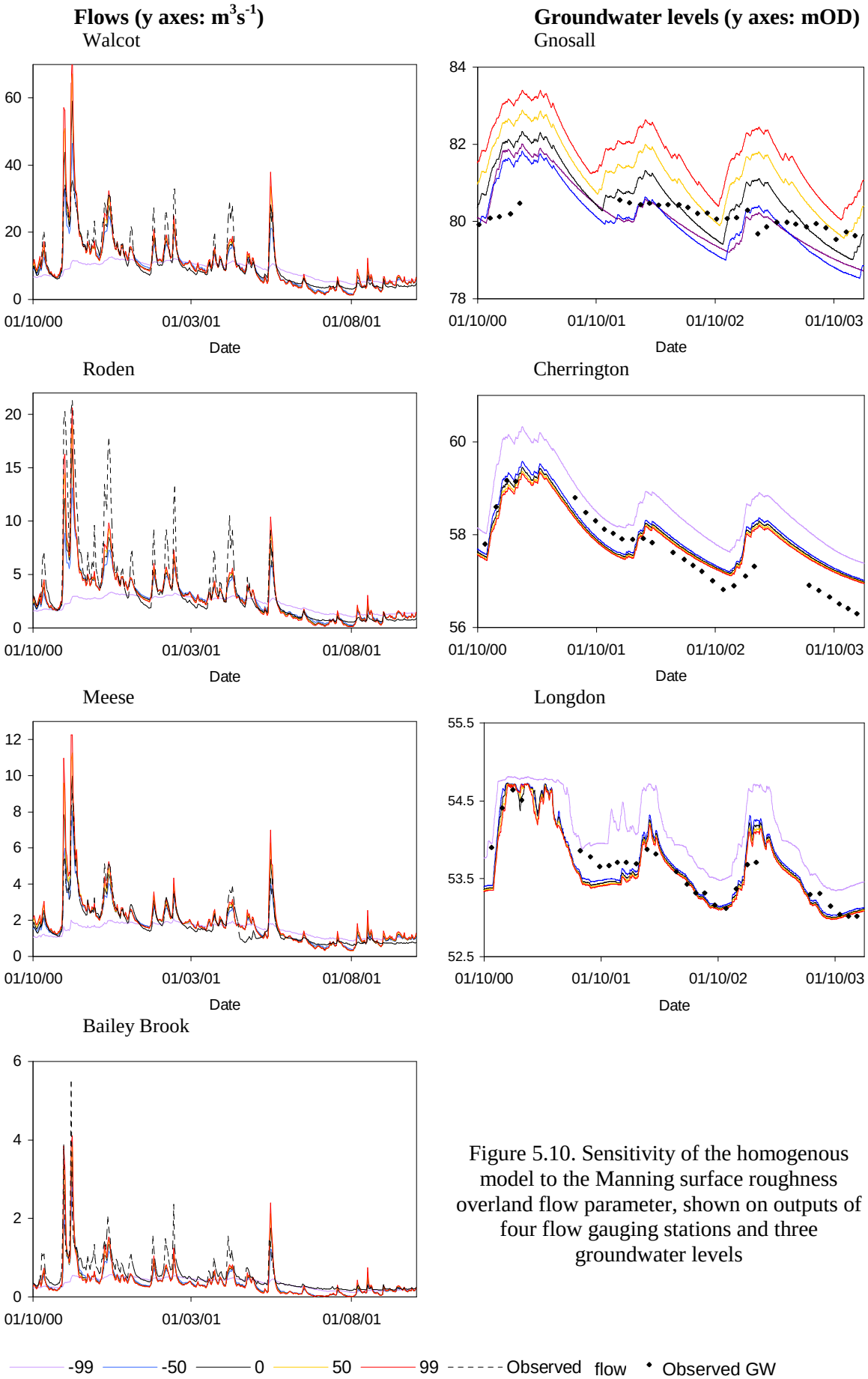


Figure 5.10. Sensitivity of the homogenous model to the Manning surface roughness overland flow parameter, shown on outputs of four flow gauging stations and three groundwater levels

Manning's M typically ranges from values of 100 (in stream flow without any surface resistance) to 10, with DHI-WE, (2005) suggesting that lower values should be used for overland flow within the catchment. The 0.005 value at the -99% perturbation is therefore indicative of an extremely rough surface where overland flow is near prohibited (with water likely to infiltrate to the sub-surface). The smooth hydrographs shown with this perturbation are therefore expected to result from baseflow with little or no overland flow runoff peaks. The corresponding increase in general groundwater levels (Figure 5.10) at the -99% perturbation reflect the corresponding increase in volume of water entering the sub-surface, as expected.

It is perhaps unexpected to see the variability in simulated groundwater levels at Gnosall to changes in Manning's M, that shows the opposite response to the other two boreholes. However, as introduced previously, Gnosall is close to the catchment boundary and the simulated response in the calibration was not good. Here it is expected that the response in Figures 5.9 and 5.10 are different to the other groundwater level sites for the same reason.

5.4.3. Unsaturated zone

Figure 5.11 shows the output flow and groundwater level variation in response to the perturbation of five unsaturated zone parameters. It is evident that the model is to varying extents sensitive to all the parameter perturbations assessed, especially the soil water content at saturated (SWC Sat) and field capacity (SWC Fc). The summary plots also suggest that the same pattern of response is seen for all sites, with the only noticeable difference between river flow and groundwater levels being the way the model responds to variation of the infiltration rate (explained later in this section).

Figure 5.11 also shows that SWC Sat and Fc could not be perturbed by the full range (as also shown in Table 5.3) as in some cases the perturbed values resulted in run-error as logically, the SWC Sat must always be higher than the SWC Fc. As a result the plots in Figures 5.12 and 5.13 display the river flow and groundwater levels for the narrower assessed range of -10 to 99% perturbation, and -50 to 20% perturbation, respectively.

Figures 5.11 and 5.12 suggest that the gauging stations all show the same variation in river flows to the same perturbations in SWC Sat, as well as consistent variation throughout the winter and summer months of the 2000-2001 year. A substantial downward or upward shift in the total hydrograph occurs as the SWC Sat value is decreased or increased, respectively. The peak flow shape and magnitude are not affected by changes to SWC Sat, but the volumes of water simulated as baseflow do vary considerably. This result is to be expected as increasing maximum water content of the soil means that a larger volume can be held within it and thus available as a baseflow source. As is also expected, the general response in the variation of groundwater levels with perturbation of SWC Sat results in increased groundwater levels (closer to the surface) when the SWC Sat is set at higher percentages, again a result of more water able to be stored and held within the sub-surface.

The SWC Fc appears to be a similarly sensitive parameter in the model, with the inverse relationship to the SWC Sat parameter (apparent in Figure 5.11). This is also shown in the plots for each gauging station and borehole in Figure 5.13. As the value of SWC Fc is decreased there is an increase in the total volume of simulated river flow, at all assessed gauging stations. This result is also expected as by decreasing the value less water is held in the soil before vertical flow begins to drain it, allowing for sub-surface flow. Both the SWC Sat and SWC Fc parameter variations result in considerable changes to the simulated flows and groundwater levels. Comparatively, the output plots in Figure 5.14 and summary plots in Figure 5.11 indicate that the model displays a smaller range in sensitivity to SWC field wilting point parameter variation, in comparison to the other UZ parameters.

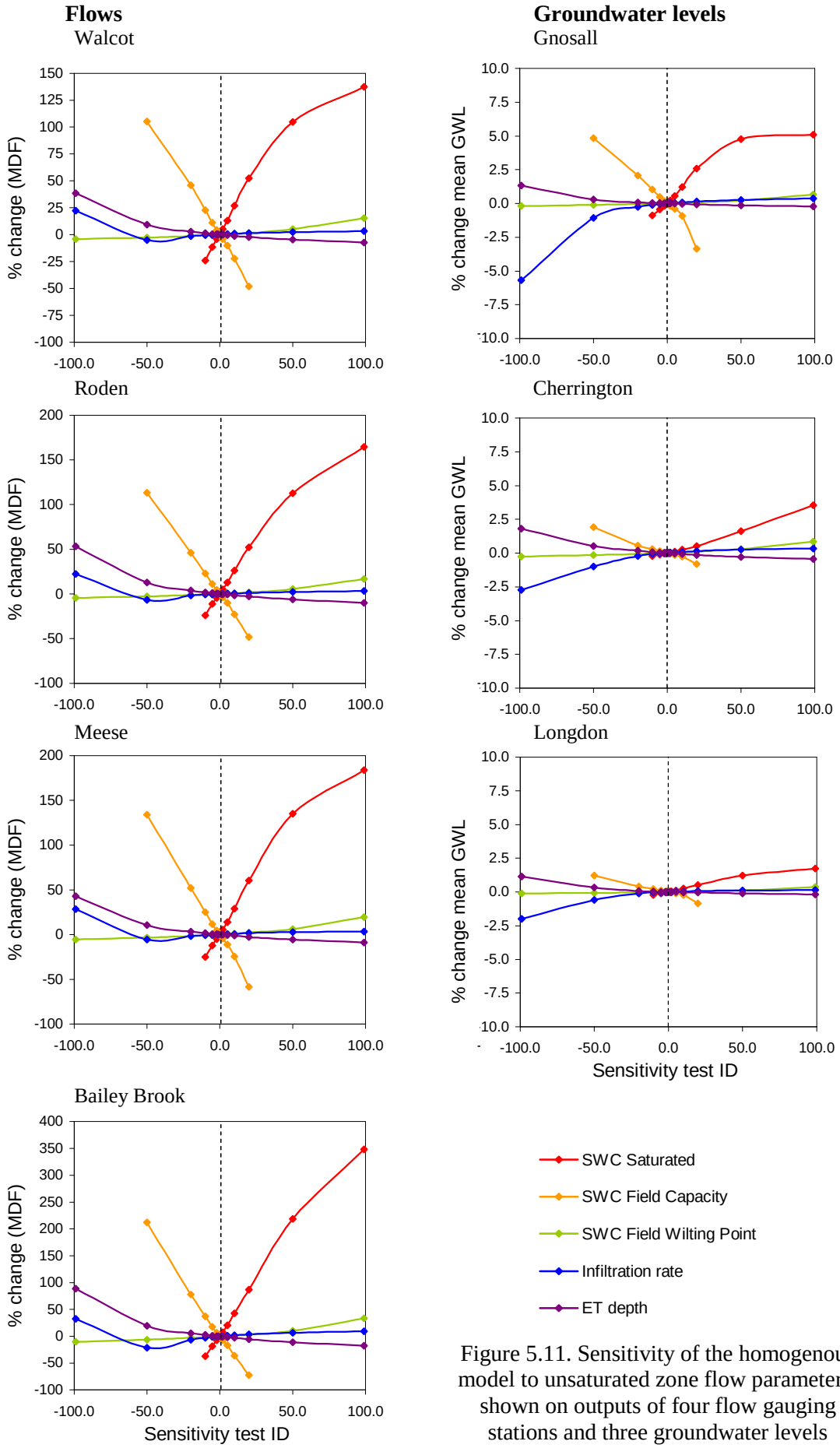


Figure 5.11. Sensitivity of the homogenous model to unsaturated zone flow parameters, shown on outputs of four flow gauging stations and three groundwater levels

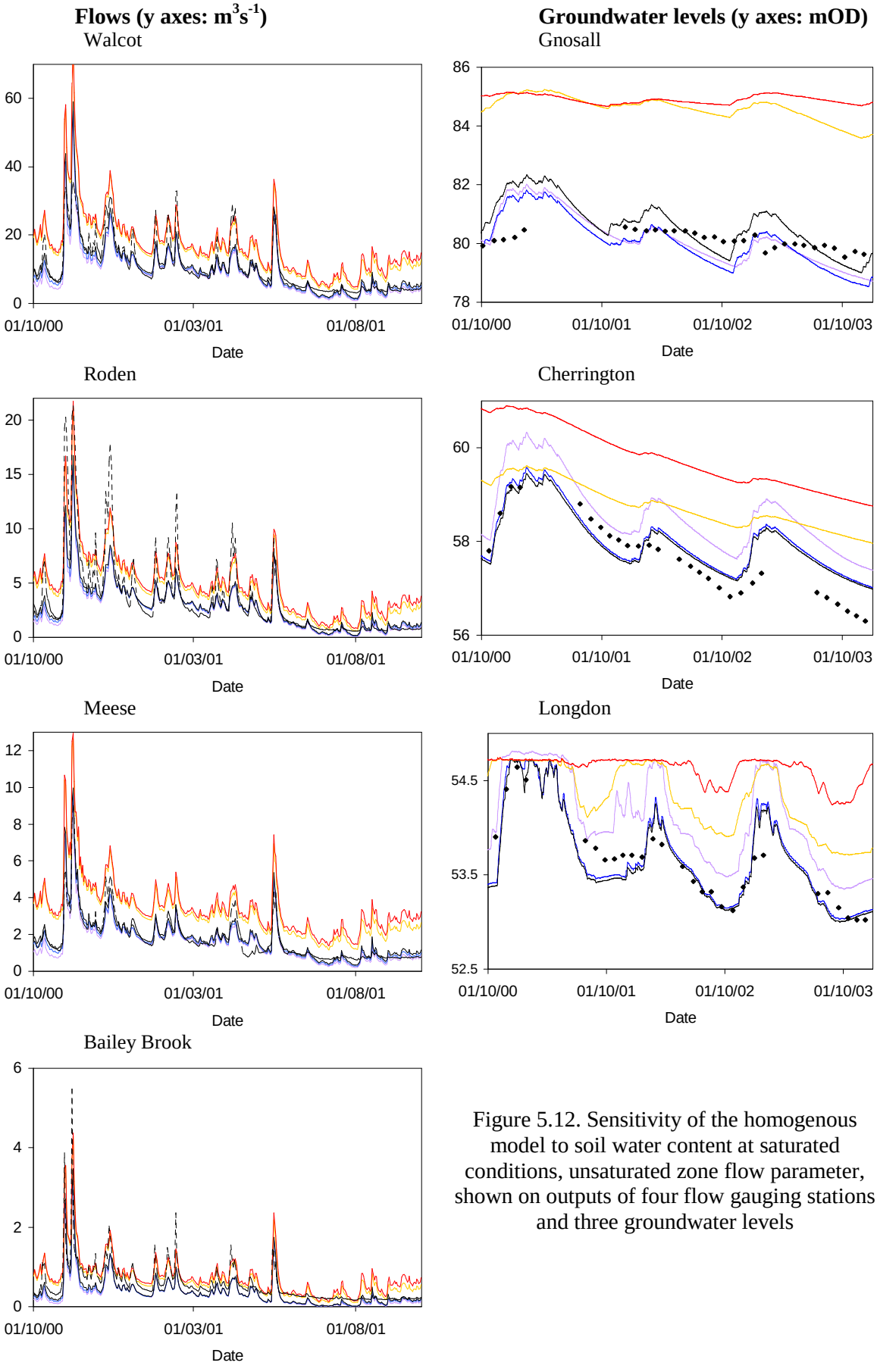


Figure 5.12. Sensitivity of the homogenous model to soil water content at saturated conditions, unsaturated zone flow parameter, shown on outputs of four flow gauging stations and three groundwater levels

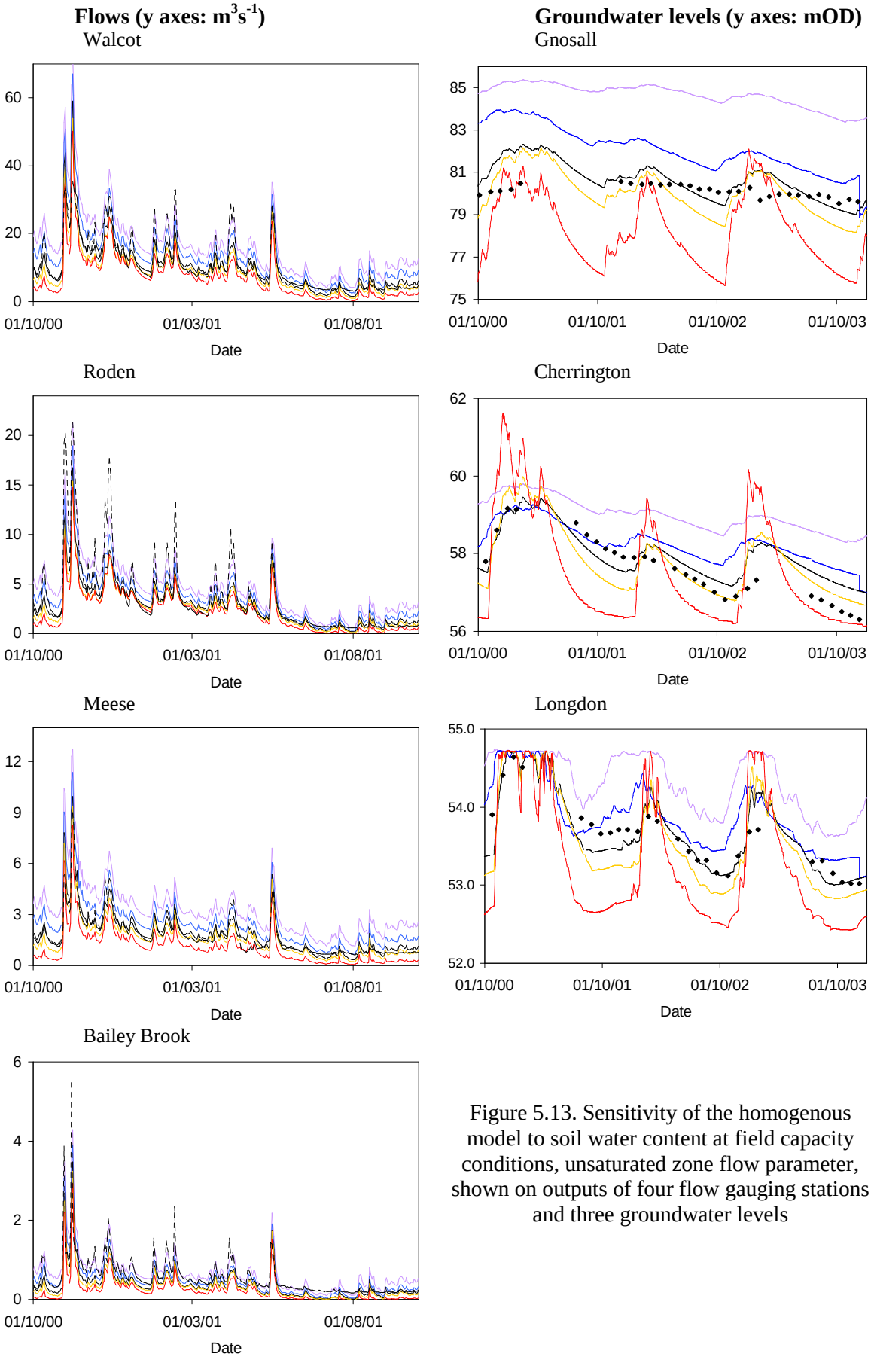


Figure 5.13. Sensitivity of the homogenous model to soil water content at field capacity conditions, unsaturated zone flow parameter, shown on outputs of four flow gauging stations and three groundwater levels

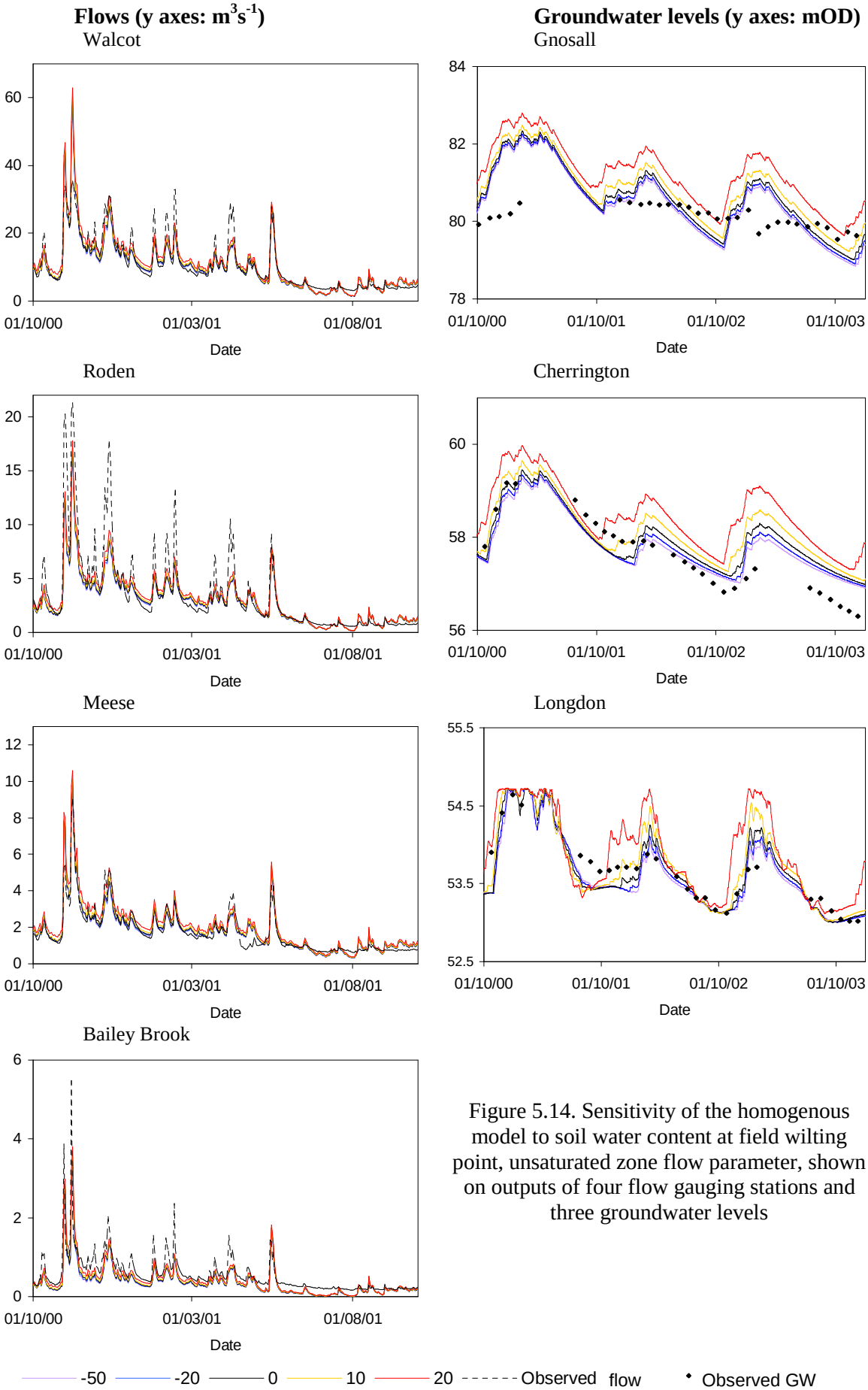


Figure 5.14. Sensitivity of the homogenous model to soil water content at field wilting point, unsaturated zone flow parameter, shown on outputs of four flow gauging stations and three groundwater levels

The results of the infiltration and ET depth parameter variations are shown in Figures 5.11 and 5.15. They suggest that an increase in infiltration rate (99% perturbation resulting in a value of 1.22 e-7 m s^{-1}, Table 5.3), there is a corresponding decrease in simulated flow peaks and the hydrograph appears less flashy, as water moves more rapidly into the soil and is not available as runoff and overland flow. Simultaneously, the ground water levels are higher when simulated with the fastest infiltration rates, confirming the model operates as should be expected.

Figure 5.11 also shows the equivalent increase in the percentage of mean daily flow in all gauging stations when the model is set with slower infiltration rates (6.12 e-10 m s^{-1} at a -99% perturbation from initial calibration). With water moving more slowly into the sub-surface (also confirmed with lower groundwater levels, Figures 5.11 and 5.15) more is available for surface runoff, as highlighted in the flashy nature of the hydrographs, especially in winter months. It is also interesting to note that seasonal variation is apparent in these hydrographs as the slower infiltration rates do not have the same effect on summer flows when baseflow typically dominates. In this case, baseflows are reduced as water is not entering the sub-surface at the same rate, and as there are characteristically fewer storm events causing flashy runoff in this season. The simulated flows during the summer at the -99% perturbation are therefore low when compared to the initial calibration that used an infiltration rate of 6.12 e-8 m s^{-1}.

This seasonal response at the -99% perturbation highlights the crude nature of the summary figures that use the mean percentage variation from mean flow or groundwater level from the initial manual calibration ‘run 0’. These plots do not include any inter annual variability and thus highlight the benefit of additionally showing the actual hydrographs of changes to river flow and groundwater levels.

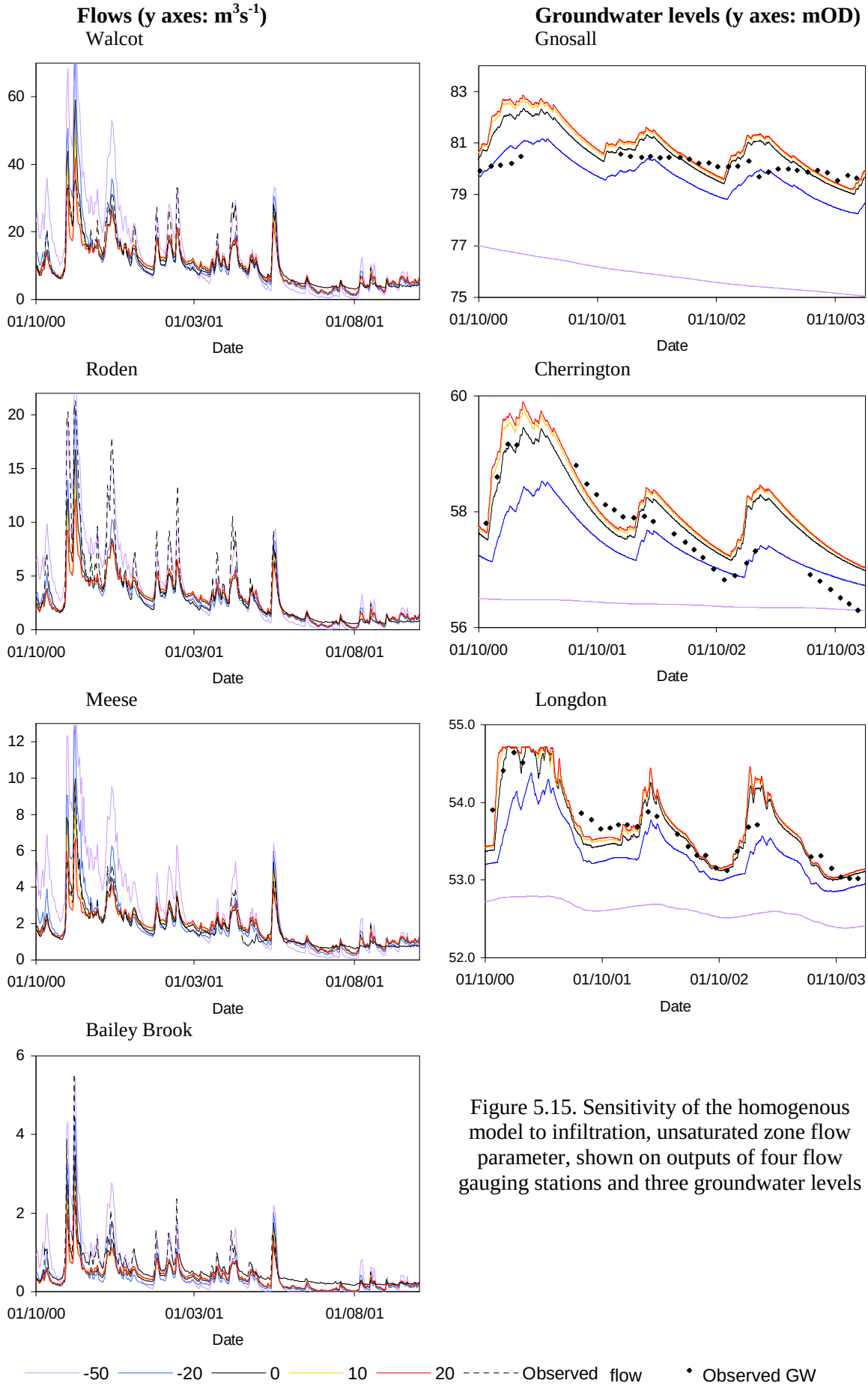


Figure 5.15. Sensitivity of the homogenous model to infiltration, unsaturated zone flow parameter, shown on outputs of four flow gauging stations and three groundwater levels

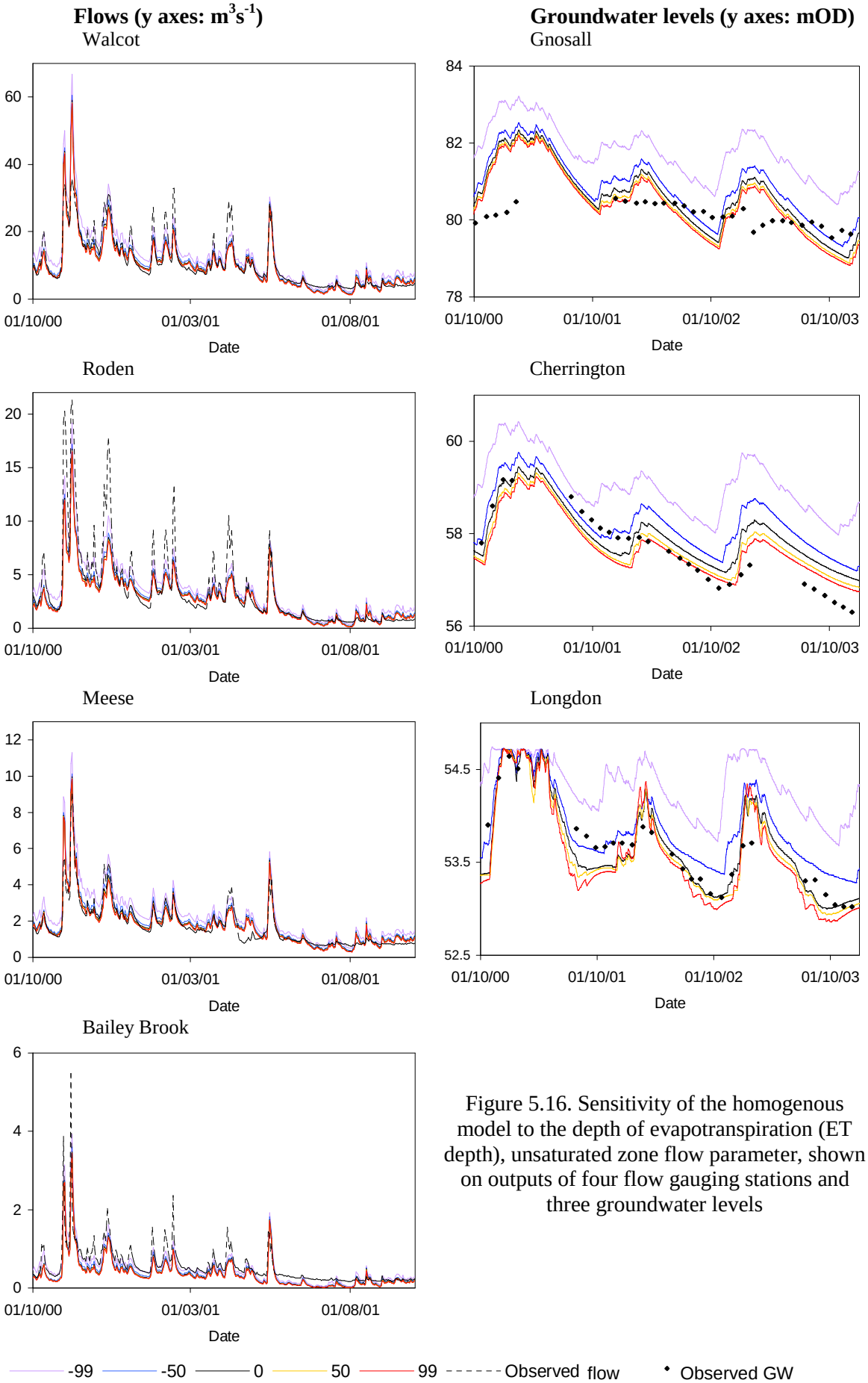


Figure 5.16. Sensitivity of the homogenous model to the depth of evapotranspiration (ET depth), unsaturated zone flow parameter, shown on outputs of four flow gauging stations and three groundwater levels

The final UZ parameter that was included in the sensitivity analyses was the depth of the ET layer. Figure 5.11 indicates that the model is less sensitive to this parameter when compared to the other UZ parameters such as SWC Sat and SWC Fc, with respect to percentage mean differences from the initial calibration. The individual site plots for flow and groundwater (Figure 5.16) do however show more specifically that the parameter has some influence on model outputs. As anticipated, the smaller the depth of ET layer (e.g. at 1m when set at -50% perturbation from initial calibration), the higher the general flow and higher the groundwater levels. This is expected as plants and vegetation can extract less water for evapotranspiration when the limit is at shallower depths, therefore there is more water left within the system available as stored groundwater or as river flow.

5.4.4. Saturated zone

Results of the sensitivity tests to saturated zone parameters indicate that the model is sensitive to the horizontal hydraulic conductivity parameter within the range of perturbations considered (Figure 5.17). In contrast, variation of the other three saturated zone parameters (vertical hydraulic conductivity, specific yield and the storage coefficient) show negligible or no impact on the percentage changes to mean daily flow or groundwater levels.

Logical response of the model to variation of hydraulic conductivity is further shown in Figure 5.18. Reduction in the rate of horizontal hydraulic conductivity to the lower values tested (7.5 e-7 m s^{-1} at -99% parameter perturbation) results in the lateral movement of water through the ground being slower. The proportion of baseflow shown in the hydrographs at all four gauging stations is therefore reduced (up to ~50% reduction in MDF for all flow sites, Figure 5.17). Accordingly, the groundwater plots suggest that with slower horizontal hydraulic conductivity, the storage of water within the saturated zone is higher, with higher groundwater levels simulated. As the parameter is altered to higher rates and lateral movement of water increases (e.g. 1.49 e-4 m s^{-1} at 99% increase from the initial manual calibration) the groundwater levels are simulated to fall to lower levels as less water is stored.

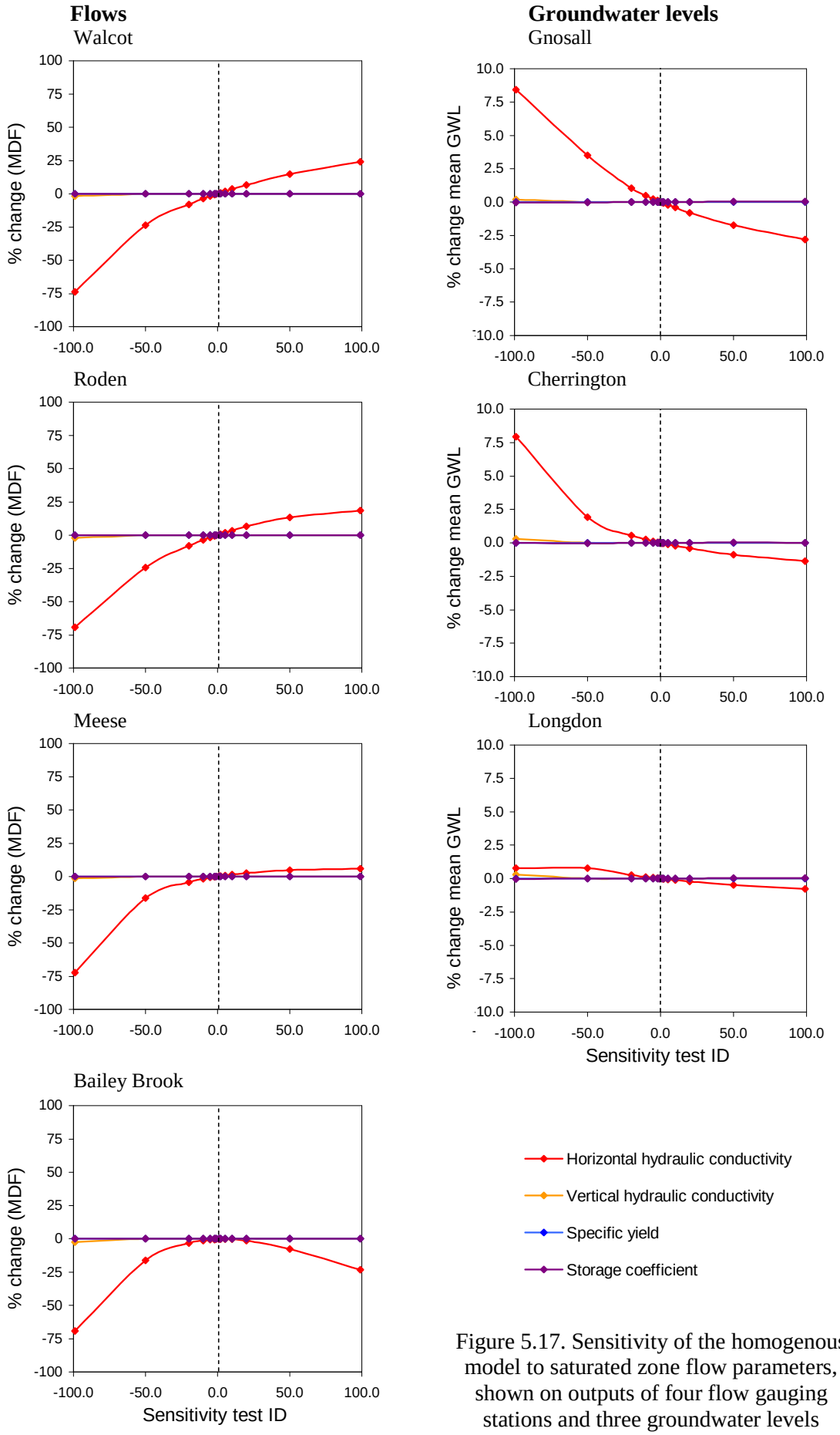


Figure 5.17. Sensitivity of the homogenous model to saturated zone flow parameters, shown on outputs of four flow gauging stations and three groundwater levels

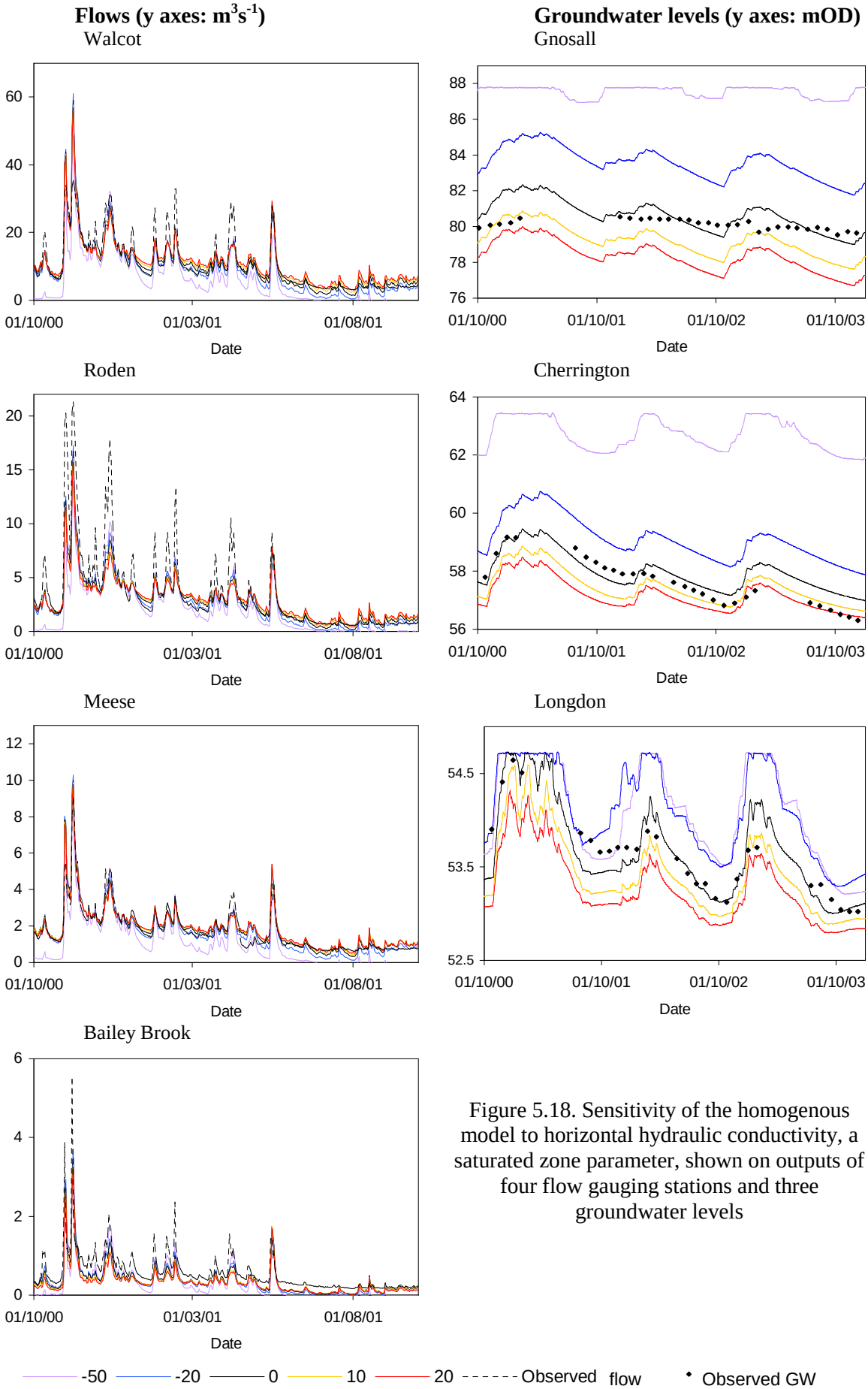


Figure 5.18. Sensitivity of the homogenous model to horizontal hydraulic conductivity, a saturated zone parameter, shown on outputs of four flow gauging stations and three groundwater levels

Figure 5.18 suggests that the 7.5 e-5 m s^{-1} value used to represent the catchment mean is generally well balanced, with increases and decreases in the value shown to considerably alter the currently well simulated baseflows and groundwater levels (in some cases by metres).

It could be expected that the model outputs may have resulted in more variation than is displayed for perturbations to the vertical hydraulic conductivity parameter if the set up of the model included more than one saturated zone layer. However, with only one layer represented in this model, the response is fair as no water is moving vertically to pass through to secondary or tertiary geological layers. Similarly this may also be the case for the specific yield and storage coefficient parameters, Figure 5.17.

5.4.5. Review of parameter sensitivity in the homogenous model

The sensitivity analysis has in general highlighted that the perturbation of the 12 model parameters result in differing responses to the magnitude and shape in simulation of flows and groundwater levels. A review is now given relating to the objectives of the sensitivity analysis that were first described in Section 5.4.

1) To undertake a model response test to the variation of parameters to assess the operation of the model, and check whether the perturbation of the parameters resulted in the expected, logical responses to simulated hydrographs and groundwater levels.

Using the sensitivity analysis as a model response test has proved beneficial in confirming the model parameters, when adjusted, result in the expected responses to changes in flow and groundwater at different sites. The exception to this is the response shown at Gnosall groundwater level borehole, where sometimes unexpected responses are likely due to the proximity of the catchment boundary.

The summary percentage changes shown Figures 5.9, 5.11 and 5.17 were useful in comparing model response at different sites, albeit in a crude calculation that did not assess intra-annual changes with parameter variations. Despite this, similar responses at all gauging stations have been shown when varying each of the parameters. The plots

have also been useful in contrasting the response of flows and groundwater levels. For example, it was demonstrated that when the speed of horizontal hydraulic conductivity was decreased there was a reduction in baseflow in the hydrographs as well as a general lowering of the groundwater levels.

It must also be noted that the purpose of the sensitivity analysis was not to undertake a detailed assessment of parameter perturbation response, but rather to assess the general operation of the model and highlight the most sensitive model parameters. The full parameter ranges for each parameter were not assessed (e.g. the infiltration rate parameter could have been assessed at higher and lower values than tested), however, the aim was to compare the parameters equally, using the same percentage changes from the initially calibrated value. Similarly, this sensitivity analysis did not assess any correlation between different parameters that may exist (e.g. Sieber and Uhlenbrook, 2005), as this was not the main purpose of the study.

2) Given the initial manual calibration presented in the preceding Section 5.3, to assess whether any of the sensitivity analyses results suggested the manual calibration could have been improved.

As shown in Section 5.3, it is believed that the initial model calibration was good, considering that the calibration aimed to balance the performance of both river flows and groundwater level using a multi-criteria approach. The sensitivity analysis has graphically confirmed what was learnt about the parameter variation and model response during the process of manual calibration. It is acknowledged that some revisions could be made that may (or may not) improve model performance, however, the balance of parameter values obtained from the documented calibration was realistic, and given that it can take a long time to manually revise and re-run each calibration whilst keeping the parameter set in balance, the existing calibration has been considered as satisfactory.

3) To quantify which parameter variations the model outputs were most sensitive to, and subsequently which parameters then carried forward for an automatic calibration procedure described in Section 5.5.

Other than assessing general parameter variation response, a further aim was to assess whether the model outputs were more sensitive to any specific components or parameter variation than others, and consequently which to carry forward in automatic calibration. Table 5.13 seeks to define a classification of parameter sensitivity on output river flow and groundwater levels, between one (no sensitivity) and five (very high sensitivity) based on overall percentage ranges (maximum percentage change minus minimum percentage change) for each parameter. The percentage ranges shown in Table 5.14 and 5.15 are calculated as the mean percentage range for all four gauging stations, and seven boreholes, respectively.

Table 5.13. Criteria for classifying model sensitivity in the homogenous model

ID	% range in mean daily flow	% range in groundwater levels	Classification
1	0.0	0.0	None
2	0.1 – 5.0	0.1 – 1.0	Low
3	5.1 – 25.0	1.1 – 2.5	Medium
4	25.1 – 100.0	2.6– 5.0	High
5	100.1 +	5.1 +	V. High

Table 5.14. Summary % ranges in sensitivity for each gauging station and parameter assessed

% Ranges to flow	Walcot	Roden	Meese	Bailey Brook	Average % range	ID	Selection
Initial water depth (OLF)	0.0	0.0	0.0	0.0	0.0	**1**	
Specific yield (SZ)	0.0	0.0	0.0	0.0	0.0	**1**	
storage coefficient (SZ)	0.1	0.2	0.1	0.1	0.1	**2**	
VK (SZ)	1.7	2.2	1.5	2.6	2.0	**2**	
Detention storage (OLF)	2.1	2.6	2.3	5.9	3.2	**2**	
SWC Fwp (UZ)	19.7	21.3	25.2	43.0	27.3	**3**	✓
Infiltration rate (UZ)	14.8	38.2	1.1	77.8	33.0	**4**	✓
Manning (OLF)	27.1	28.4	33.9	53.4	35.7	**4**	✓
ET Depth (UZ)	46.1	63.4	51.5	106.5	66.9	**4**	✓
HK (SZ)	97.8	87.8	78.0	69.1	83.2	**4**	✓
SWC FC (UZ)	153.4	161.3	192.0	284.2	197.7	**5**	✓
SWC Sat (UZ)	161.5	188.3	208.8	385.5	236.0	**5**	✓

Table 5.15. Summary % ranges in sensitivity for each borehole and parameter assessed

% Ranges to GWL	Gnosall	Cherrington	Longdon	Average % range	ID	Selection
Initial water depth (OLF)	0.0	0.0	0.0	0.0	**1**	
Specific yield (SZ)	0.0	0.0	0.0	0.0	**1**	
storage coefficient (SZ)	0.0	0.0	0.0	0.0	**1**	
VK (SZ)	0.2	0.3	0.3	0.3	**2**	
Detention storage (OLF)	0.1	0.0	0.0	0.1	**2**	
SWC Fwp (UZ)	0.9	1.1	1.4	1.1	**3**	✓
Infiltration rate (UZ)	6.1	3.0	2.2	3.8	**4**	✓
Manning (OLF)	2.0	1.2	0.9	1.4	**3**	✓
ET Depth (UZ)	1.5	2.2	1.4	2.7	**4**	✓
HK (SZ)	11.2	9.3	1.6	7.4	**5**	✓
SWC FC (UZ)	8.9	2.7	2.0	4.3	**4**	✓
SWC Sat (UZ)	6.0	3.7	1.9	3.9	**4**	✓

The percentage range threshold is much higher for flows than groundwater as the starting units are smaller; therefore any given change to flows from the manual calibration (control) appears larger. To enable a comparative selection of parameters, the two output measures had to be classified separately. For example, Table 5.15 indicates that a 3.8 % overall change in groundwater levels from the perturbation of the infiltration rate parameter results in a considerably large change to simulated levels. Comparatively, the mean 33% change to mean river flow (Table 5.14) shows a similar magnitude of output sensitivity and is classified into the same 'high' sensitivity.

The threshold classification in Table 5.13 has been used to quantify the parameter sensitivity so as to select parameters to use within automatic calibration. Any parameter that was classified in band 3 (medium) and above, for both flow and groundwater were selected. Tables 5.14 and 5.15 show that all of the unsaturated zone parameters are included within classes three to five (medium to very highly sensitive). In contrast, only one of the overland flow parameters (the Manning's roughness coefficient, M) has been classified as sensitive to model outputs, whilst only the horizontal hydraulic conductivity parameter from the saturated zone was selected. In total, seven parameters were selected for use in the following automatic calibration.

5.5. Automatic Calibration

This section details the methodology and results of two automatic calibration methods. The first automatic calibration optimises using only observed river flow at the basin outlet, and the second employs a comprehensive multi-location and multi-criteria approach. The approach used for both automatic calibration methods have the same simulation specifications, model parameters and parameter optimisation criteria using Monte-Carlo sampling of the seven sensitive parameters (described in Sections 5.5.1 and 5.5.2 and 5.5.4 respectively). The difference between the two automatic calibration tests is that the objective functions (Section 5.4.3) differ between set ups. The purpose of undertaking these automatic calibrations are:

1) To test whether the automatic calibration methodology (that is being more widely adopted as a preferred method to calibrate distributed models as reviewed in Section 2.5.3.2) results in an improved simulation ability of the homogenous Tern hydrological model with regard to the same river flow and groundwater level statistics used in the manual calibration in Section 5.2.

2) To compare the complexity of the automatic calibration methods, and whether by including more rigorous objective functions (in the second test) the ability of the model simulations of river flow and groundwater levels are improved.

3) To acknowledge and quantify equifinality within the parameter space. This is undertaken by recognition and selection of a suite of calibrated models for each automatic calibration providing a total of 20 calibrated models with different parameter values.

5.5.1. Methodology

As introduced in Chapter 1, hydrological models developed in MIKE SHE can be dynamically coupled to an automatic calibration component, Autocal, within the MIKE ZERO framework. Autocal uses commonly used optimisation methods such as the Shuffled Complex Evolution method (SCE) (Duan et al., 1993), previously described in Section 2.6.3.3.

Autocal requires an already existing hydrological model, in this case the manually calibrated homogenous Tern model to be specified and used as a template from which to undertake the automatic calibration. For means of comparison, the automatic calibration evaluation period has been kept the same as that used in the manual calibration (01/01/2000 – 31/12/2003), with the same warm up period (beginning 01/06/1998) and the same iterative hotstart initial conditions specified.

The seven calibration parameters have been included in the automatic calibration that were selected as a result of the parameter sensitivity analysis in Section 5.4. The method of automatic calibration adopts the parameter optimisation approach that seeks to find the mathematical optimum for each parameter based on given objective functions. The method seeks to find these optima by running simulations a number of times until given stopping criteria are met. Table 5.16 summarises the specified parameter ranges, the lower and upper bounds, as well as the initial value (derived from the manually calibrated model for each parameter). In each simulation of the automatic calibration, the parameter set is determined by Monte-Carlo sampling, for a fair and non-biased simulation. The parameter ranges have been selected based on realistic and feasible bounds defined by suggested values within the literature, and attempt to be both wide enough to capture the intra-basin variation for each parameter whilst at the same time limited as best possible so that they are still realistic.

Table 5.16. Initial, lower and upper parameter values used in automatic calibration for the homogenous model

Parameter	Type	Initial value	Lower bound	Upper bound	Transformation	Condition
Manning (OLF)	Variable	0.5	0.4	1.5	Real	
SWC SAT	Variable	0.5	0.3	0.55	Real	
SWC FC	Dependent	-	-	-	Real	((SWC SAT-SWC FWP)*0.8+SWC FWP)
SWC FWP	Variable	0.18	0.05	0.2	Real	
Infiltration rate	Variable	6.12e-8	3e-8	9e-8	Logarithmic	
Depth to ET surface	Variable	2	0.5	2.5	Real	
Horizontal hydraulic conductivity	Variable	7.5e-5	1e-6	9e-5	Logarithmic	

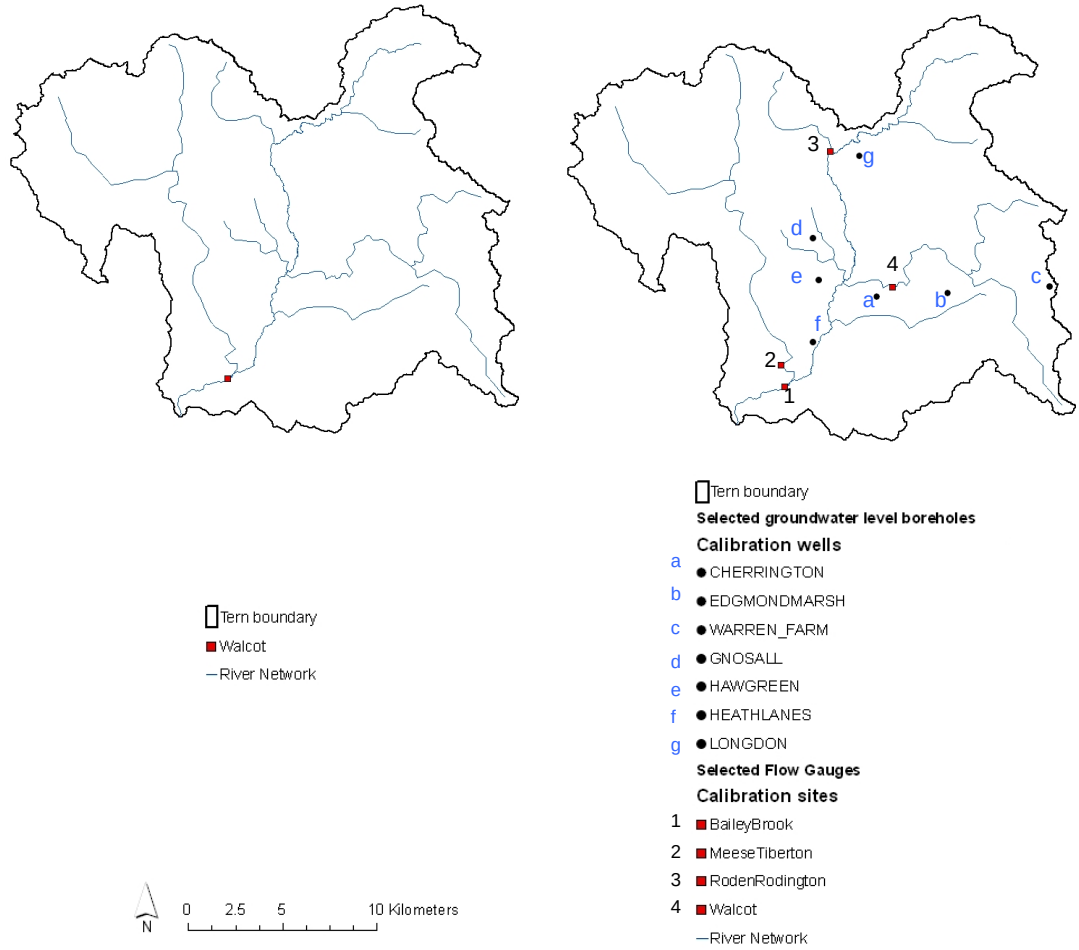
Table 5.16 shows the six variable parameters with one dependent parameter, the soil water content at field capacity. This is fixed to always be a fraction of 0.8 added to SWC-FWP of the interval between SWC-Sat and SWC-Fwp. It has been necessary to

limit it in this manner as the nature of the Monte-Carlo approach to selecting parameter values in each simulation could result in SWC-Fc values that are greater than the given SWC-SAT value. In reality and modelling terms this is not a possible option and causes the model simulation to crash. For this reason, the SWC-Fwp lower bound is set as potentially quite low (0.05) to allow for a larger variation in the more sensitive SWC-Fc parameter.

As previously introduced, automatic calibration has been undertaken for two methods, one that uses only one output measure, the RMSE of flow at the basin outlet, Walcot, to automatically calibrate the homogenous model, and the other that adopts a more rigorous procedure. The second uses a multi-criteria and multi-location approach with two objective functions that are weighted equally. The aim is to assess overall performance and not to assign any particular river flow or groundwater level as more important. One objective function sums the RMSE of river flow (the same four gauging stations used in manual calibration), and the other sums the RMSE of groundwater level boreholes (at the same seven locations used in the manual calibration) as shown in Figure 5.19. A similar method of using two objective functions, the RMSE of a set of gauging stations and the RMSE of a set of groundwater level boreholes was employed in Madsen (2003) to improve the manual calibration of the model of Refsgaard (1997) for the Karup catchment, Denmark.

It is noted that this Autocal method (also used by Madsen, (2003)) only uses one statistic in the objective function, the RMSE. Elsewhere in the research an aim has been to calibrate models based on a range of statistics. However, the version of Autocal that was used (within MIKE ZERO, 2008) did not include any of other statistics used in the thesis such as Nash-Sutcliffe, NSE or the correlation coefficient, R. Despite this, the other statistics and performance measures (Table 5.1) have been calculated retrospectively outside of Autocal, to enable a fuller discussion of model performance.

Both sets of automatic calibrations use almost exactly the same algorithmic parameters within the Shuffled Complex Evolution method of parameter optimisation. Table 5.17 summarises these parameters and the stopping criteria which are in accordance with the Autocal recommended values (DHI-WE, 2005). The autocal simulation stops when any one of the three criteria is reached.



1 output measure:
RMSE river flow at Walcot

2 objective functions:
Summed RMSE 4 gauging stations
Summed RMSE 7 Groundwater level boreholes

Figure 5.19. Comparative objective functions for the different set-ups of homogenous Tern model automatic calibration

Table 5.17. Summary of algorithmic parameters and stopping criteria used in the automatic calibration of the homogenous model (where n is the number of variable parameters, in this case, six)

Algorithmic parameter / *stopping criteria	**Recommended value**	**Value used**
Number of complexes	-	3
Number of points in a complex	2n+1	13
Number of points in a sub-complex	n+1	7
Number of evaluation steps by each complex before shuffling	2n+1	3
*Maximum number of model evaluations	-	1000 & 4000
*Number of loops of convergence	-	3
*Minimum relative change in objective function value	-	0.01

5.5.2. Automatic calibration and testing at the basin outlet (Walcot)

Using the methodology outlined in the previous section, the automatic calibration of the Tern homogenous model using the output measure of RMSE of flow at the basin outlet, Walcot (shown in Figure 5.20), resulted in convergence criteria being met after 586(+1 initial) simulations of the model using different parameter sets. The optimal RMSE value for flow at Walcot that was achieved was 2.574 m^3s^{-1} (RMSEP = 0.349). A selection of statistically calibrated models and discussion of the variation within the calibrated parameter space is now discussed, followed by the presentation of calibrated flows and groundwater levels, with their validation and testing at uncalibrated sites to follow.

5.5.2.1. Selection of calibrated models and parameter variation

Figure 5.20 presents the optimisation plot for the automatic calibration of the homogenous model using the RMSE at Walcot as the output measure by which to assess model performance. The plot identifies the consistent improvement in the RMSE from simulation ~300 onward. After this point, the range of improvement of the RMSE is small, with 184 parameter sets/simulations being contained within 2% of the optimum value, 2.574. When considering the earlier discussed equifinality of parameter spaces within a hydrological model, (Section 2.6.3), the mathematical optimum set of parameters do not necessarily result in the only statistically viable parameterisation and only calibrated model. Instead, the optimal parameter set may result in one of many statistically viable models from the region close to the optimum RMSE. Considering a relatively small threshold of a 2% allowance from the optimal RMSE (in this case the region where the RMSE is between 2.574 and 2.625 m^3s^{-1} results in 184 different parameter sets that can be considered within the optimum calibrated range. Therefore there are a suite of models that are statistically 'calibrated'. The RMSE of these calibrated models are shown in Figure 5.20 (highlighted in red), which are located beneath the 2% of optimal RMSE threshold line. Additionally, the optimal calibrated model is also shown in this plot, as well as a sample of ten selected calibrated models, discussed later in this section.

In addition to the optimisation of the RMSE shown in Figure 5.20, the optimisation of the individual parameters is highlighted in Figure 5.21. These plots show the values of each parameter adopted with each model run. It is clear that all but the horizontal hydraulic conductivity and SWC Sat optimise to a small region within the allowed parameter space, with the same 184 calibrated parameter sets within the 2% threshold of optimal RMSE shown in red. It is also interesting to note that the parameters have optimised at different places within the defined parameter limits, with ET depth, SWC Fwp and the surface roughness coefficient optimising close to the lower part of the parameter space, and the other parameters converging to optimisation near the upper end of their parameter spaces.

For each of the 587 model simulations with different parameter sets, the time series of flows and groundwater levels were stored at multiple sites within the catchment that have already been subject to calibration and testing in the manual calibration and sensitivity analysis. However, it was not possible to store additional statistics such as the Nash-Sutcliffe NSE and the Correlation, R, within the Autocal results, and so to calculate these, the new parameter sets were re-simulated to define new MIKE SHE models.

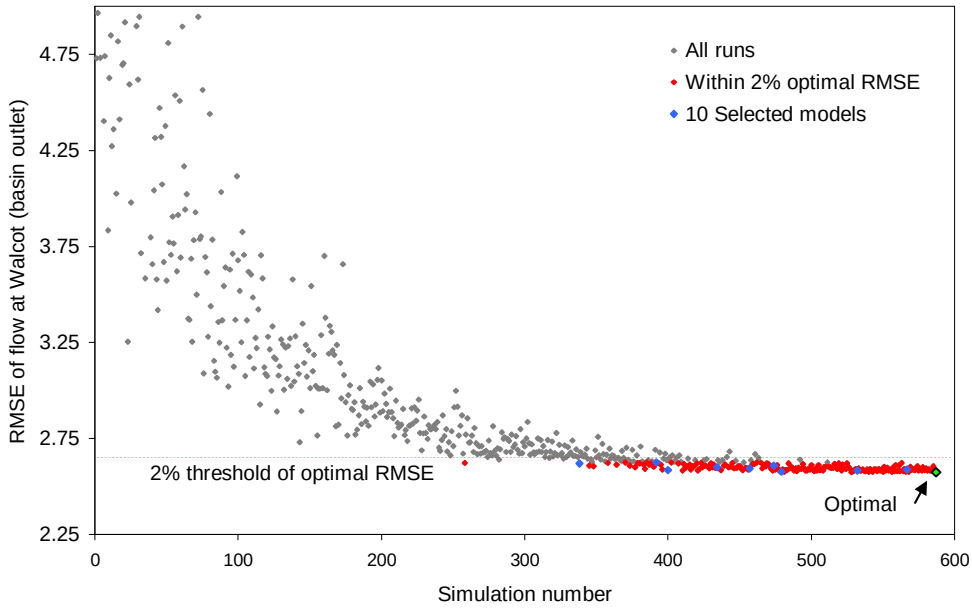


Figure 5.20. Optimisation of RMSE m^3s^{-1} at the basin outlet for the automatically calibrated homogenous Tern model

Due to logistical constraints and computational time, only a representative sample of ten calibrated models have been re-simulated through MIKE SHE. The sample of models

seek to include the variation within the calibrated range, by including the optimal model with the lowest RMSE, the model/parameter set with the RMSE at the upper limit of the 2% from optimal range, and a systematic equidistant sample in between (every 23rd model simulation from the order in which the models were simulated). The resulting ten models are highlighted within the optimisation plot, Figure 5.20, and shown to be well distributed within the calibrated range. The specific parameter values for the sampled ten models as well as the range in values for all 184 models with the RMSE within the 2% range from optimal RMSE are summarised in Table 5.18.

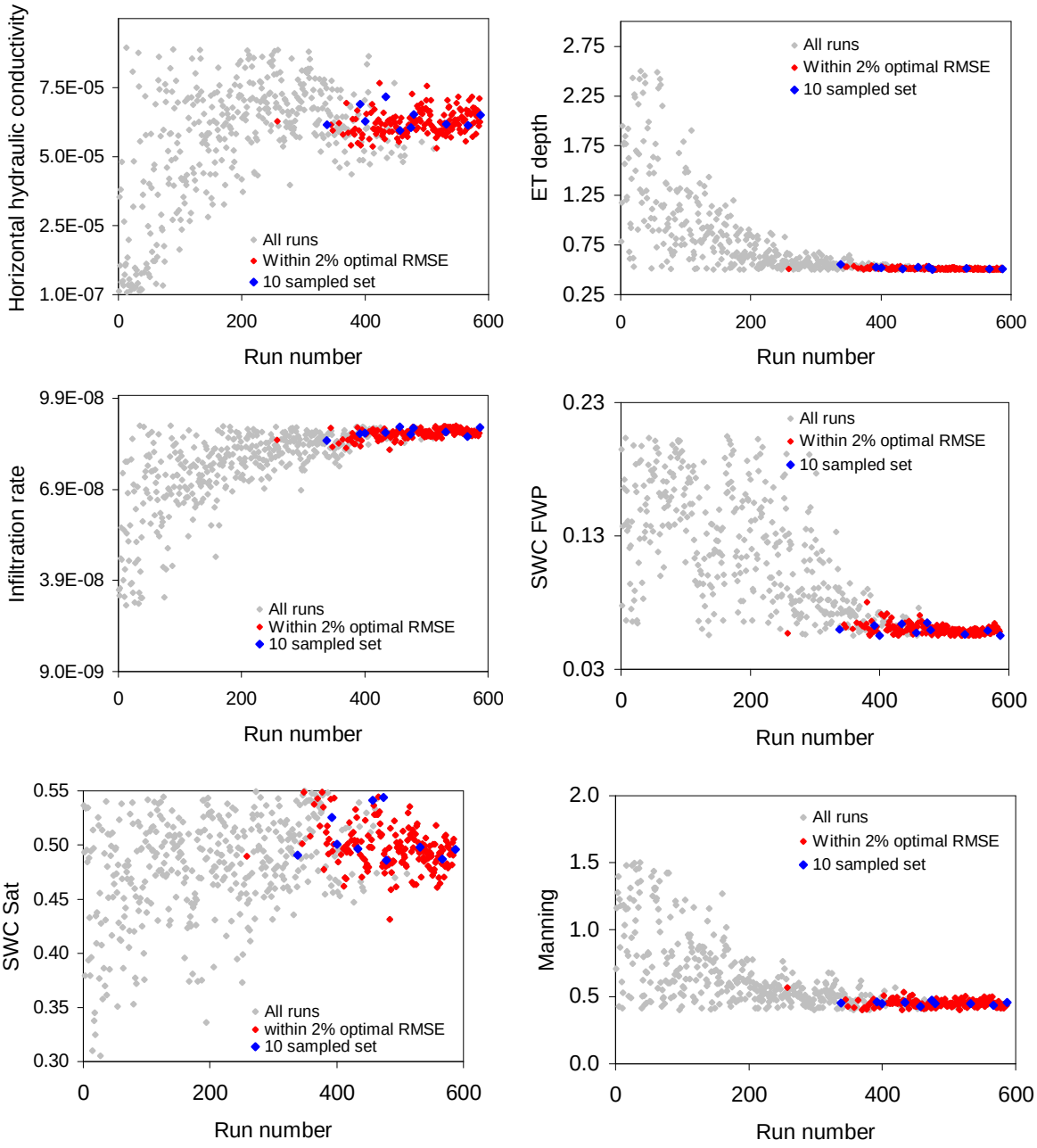


Figure 5.21. Optimisation of each parameter during automatic calibration at the basin outlet of the homogenous model. Units for y axes: SWC Sat and SWC FWP (fractions), Infiltration rate and horizontal hydraulic conductivity (m s^{-1}), ET Depth (m)

Table 5.18 indicates that the ten sampled models include the same minimum and maximum RMSE as the 184 calibrated sets, thus covering the range of RMSE values within the larger set of models. The sampled ten models do not cover the entire minimum and maximum ranges of calibrated parameter ranges but still include the majority of the variability in these parameter values.

Table 5.18. Parameter values for ten sampled models within the context of minimum and maximum parameter ranges for the homogenous model automatically calibrated at the basin outlet

ID	Run ID	Manning	SAT	FWP	Infiltration (ms^{-1})	ETdepth (m)	HK (ms^{-1})	RMSE m^3s^{-1} Flow Walcot
1	587	0.4572	0.4962	0.0501	8.941E-08	0.5078	6.504E-05	**2.574**
2	479	0.4486	0.4860	0.0545	8.921E-08	0.5012	6.540E-05	**2.576**
3	532	0.4488	0.4982	0.0511	8.804E-08	0.5131	6.189E-05	**2.582**
4	400	0.4479	0.5008	0.0502	8.757E-08	0.5200	6.290E-05	**2.586**
5	567	0.4348	0.4871	0.0539	8.639E-08	0.5107	6.131E-05	**2.589**
6	457	0.4279	0.5416	0.0523	8.966E-08	0.5254	5.955E-05	**2.595**
7	434	0.4572	0.4967	0.0594	8.777E-08	0.5067	7.171E-05	**2.600**
8	474	0.4761	0.5444	0.0600	8.730E-08	0.5234	6.065E-05	**2.608**
9	338	0.4518	0.4908	0.0547	8.522E-08	0.5541	6.155E-05	**2.619**
10	392	0.4632	0.5254	0.0578	8.744E-08	0.5265	6.909E-05	**2.625**
Parameter ranges of the 10 sampled sets (above) within 2% optimal RMSE								
Minimum		0.4279	0.486	0.05012	8.522E-08	0.5012	5.955E-05	**2.574**
Maximum		0.4761	0.5444	0.06002	8.966E-08	0.5541	7.171E-05	**2.625**
Parameter ranges of all 184 calibrated sets within 2% optimal RMSE								
Minimum		0.4006	0.4315	0.05005	8.215E-08	0.5002	5.32E-05	**2.574**
Maximum		0.5702	0.5492	0.07548	8.997E-08	0.5541	7.67E-05	**2.625**

5.5.2.2. Calibrated model results and testing/validation

Figure 5.22 plots the calibrated bounds (minimum and maximum daily flows) for the suite of ten sampled models at the four river flow gauging stations used in the manual calibration in Section 5.3. Although these bounds are derived from an automatic calibration only at the basin outlet, Walcot, the additional gauging stations are used to test and validate the automatic calibration procedure to see how well the calibrated models simulate flows internally within the catchment. Figure 5.22 also includes the observed and manually calibrated model outputs to facilitate comparisons of model performance.

The variability of the parameter ranges shown in Table 5.18 translated in to very minimal differences in calibrated flows at all gauging stations, not just Walcot for which the model has been calibrated. This confirms that there can be a multitude of parameter sets from which almost equally statistically viable models can be derived. This equifinality draws to attention the uncertainty within the calibration process.

Figure 5.22 also suggests that a model automatically-calibrated at the basin outlet is able to produce good internal river flow simulations within different sub-basins of the catchment. Upon inspection, the suite of automatically calibrated models appear to all give a very comparable, if not better simulation results when compared to the manually calibrated model, with the Meese tributary in particular appearing very well calibrated, NSE of 0.81-0.82 for the suite of ten calibrated models compared to 0.75 in the manual calibration.

The calculation of the summary score measure has also been undertaken for the suite of ten automatically calibrated models, with the mean score for all four gauging stations at 0.7725 compared to 0.695 for the manual calibration (Table 5.19). These results therefore support the quantitative statistics in suggesting the automatic calibration at the basin outlet results in general improvement at all of the river flow gauging stations.

Table 5.19 Results of summary score for river flows from automatic calibration at the basin outlet for the homogenous model

Flow	**Walcot**	**Roden**	**Meese**	**Bailey Brook**
Baseflow	0.21	0.18	0.2	0.12
Peaks	0.19	0.17	0.22	0.19
Timing	0.25	0.24	0.23	0.19
MDF	0.22	0.17	0.24	0.07
Total	**0.87**	**0.76**	**0.89**	**0.57**

Mean score **0.7725**

In addition to flows, minimum and maximum calibrated bounds of groundwater level simulations from the sampled set of ten models are shown in Figure 5.23 for the same seven boreholes used in the manual calibration (Section 5.3). The plots also show the observed level data as well as the manually calibrated outputs to enable a comparison of model performance. All groundwater levels were entirely excluded from any objective functions in this automatic calibration, and so are also used to test the calibration in addition to the internal gauging stations shown previously. Overall, a similar magnitude in calibration bounds are seen at all boreholes (i.e. all ten parameter sets simulate

groundwater within one metre variation), with the largest (but not unrealistic) range seen at Hawgreen.

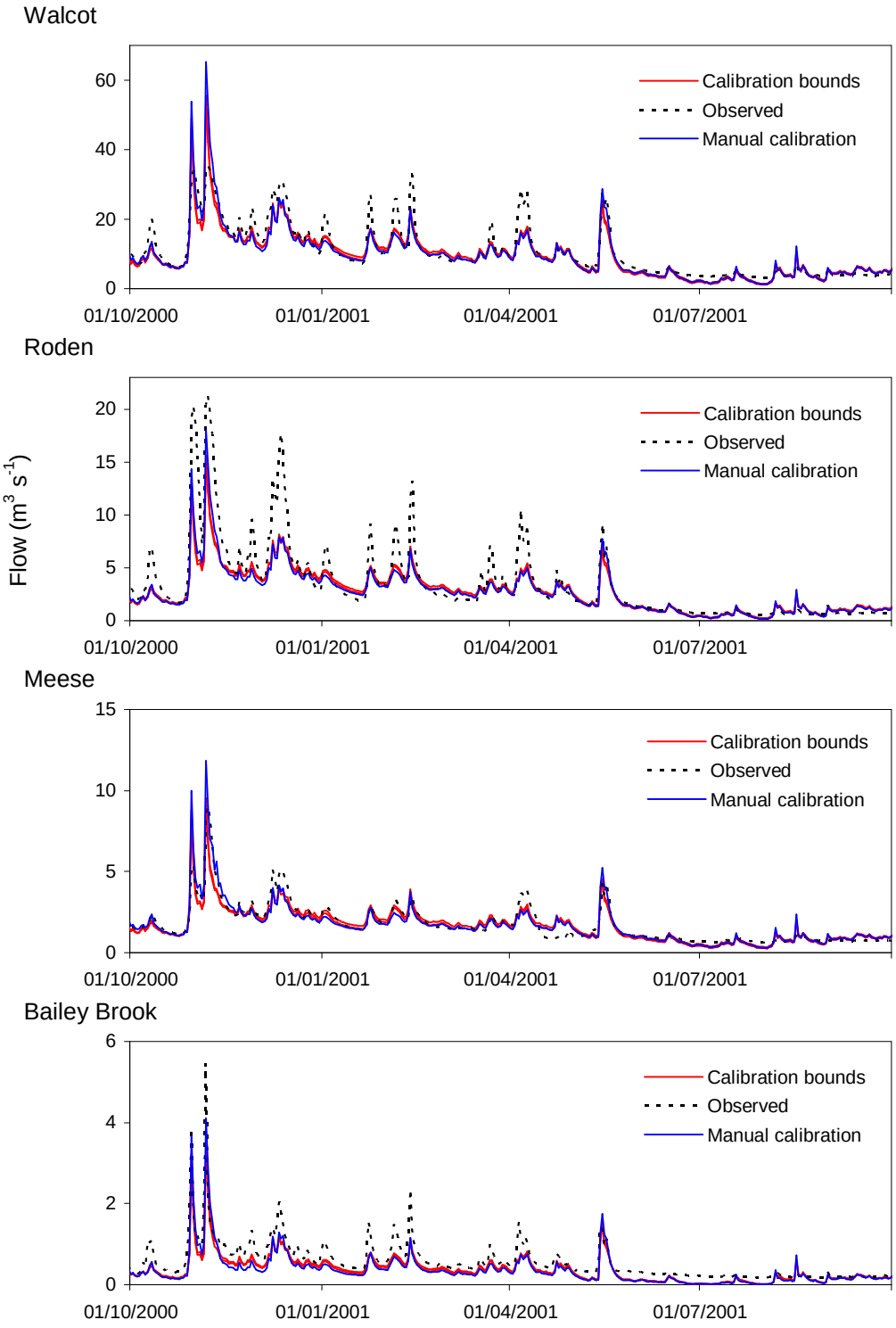


Figure 5.22. Observed and simulated calibration bounds at four gauging stations from the homogenous model, calibrated at the basin outlet. The model is calibrated at Walcot and tested at the further three internal river flow gauging stations

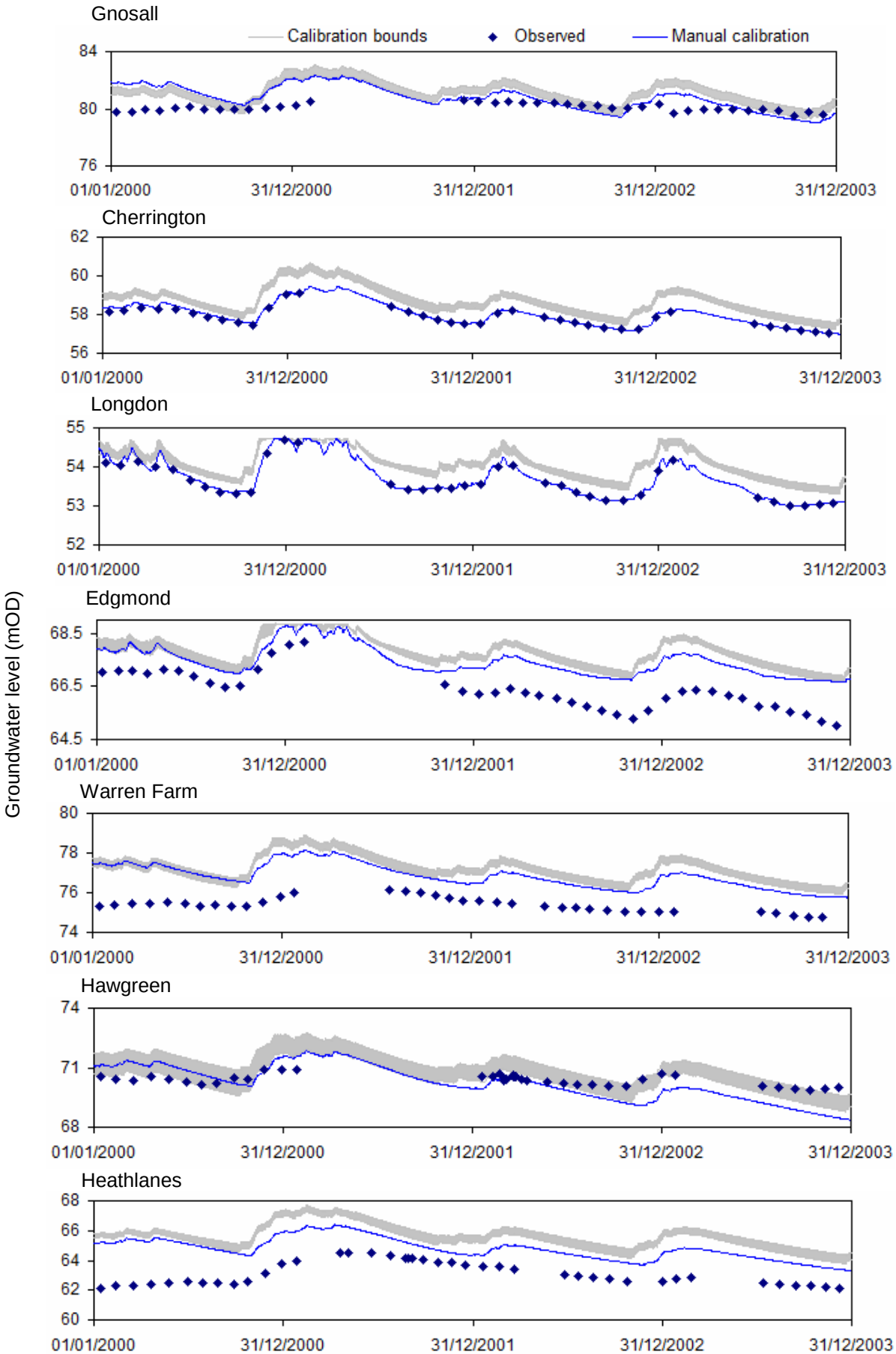


Figure 5.23. Results of calibration bounds derived by autocalibration of the homogenous model of river flow at the basin outlet, for seven groundwater level sites that were not used in the calibration

For an automatic calibration not considering groundwater levels, and only set up to optimise the RMSE of flow at the basin outlet, the results in Figure 5.23 are considered good. At most groundwater level boreholes the seasonal variations and shape of groundwater levels are simulated relatively well. However, the results in Figure 5.23 highlight that the manual calibration of groundwater levels were better in all boreholes other than Hawgreen, when compared with the observed data. In general, the mean/general level of groundwater in the automatically calibrated model appears to be simulated less well, with levels typically 1.0 – 1.5m above the observed groundwater levels.

Table 5.20 summarises key statistics from the ten selected models based on the automatic calibration at the basin outlet. For each location both the minimum and maximum values of mean river flow or groundwater level, total flow, Nash Sutcliffe NSE, Correlation R, RMSE and RMSEP for the suite of ten calibrated models are shown. Performance in the simulation of flow at the basin outlet appears improved from that of the manual calibration, with an improvement in the Nash-Sutcliffe coefficient (previously 0.74 improved to 0.78 – 0.79), correlation (previously 0.81 improved to 0.89 – 0.90), and RMSE (previously 2.87 m^3s^{-1} improved to 2.57 – 2.62 m^3s^{-1}). Likewise, an improvement in these three statistics are seen for the Meese and Bailey Brook gauging stations, with similar values to those achieved in manual calibration seen for the Roden. The ranges in the lower and upper limits (minimum and maximum) for the suite of ten sampled models for each river flow statistic are small, confirming that different parameter set calibrations can yield almost identical statistical as well as plotted river flow and groundwater levels, not just for the RMSE statistic used in calibration, but for other statistics too. It is also noted, however, that although each parameter set have different values, Table 5.18 highlighted that the actual variation within the parameter space was relatively small, likely due to the optimisation being based on just one output measure in this automatic calibration method.

Statistics for the groundwater level simulations in Table 5.20 indicate a good overall performance despite groundwater level observations not being included in the automatic calibration. The range of correlation coefficients for the suite of ten models are comparatively larger than for the simulated flows, with good correlation for all

boreholes except Gnosall. RMSE statistics are also widely ranging between boreholes and within the sampled set of models, with Longdon and Hawgreen showing the best performance.

Table 5.20. Summary performance statistics for homogenous model calibrated at basin outlet

Name	Measure	Min - max MDF (m^3s^{-1}) or MGWL (m)	Range (m^3s^{-1}) or (m)	Total flow m^3s^{-1}	NSE	R	RMSE Flow (m^3s^{-1}) GWL (m)	RMSEP
Walcot	Flow	6.60 – 6.92	0.32	7.99E+08 - 8.38E+08	0.78 – 0.79	0.89 – 0.90	2.57 – 2.62	0.35 - 0.36
Roden	Flow	1.82 – 1.90	0.08	2.30E+08 - 2.40E+08	0.71 – 0.72	0.90 – 0.91	1.26 – 1.30	0.59 - 0.61
Meese	Flow	1.15 – 1.22	0.07	1.46E+08 - 1.55E+08	0.81 – 0.82	0.90 – 0.91	0.35 – 0.36	0.29 - 0.30
Bailey Brook	Flow	0.24 – 0.25	0.01	3.05E+07 - 3.17E+07	0.58 – 0.62	0.90 – 0.90	0.20 – 0.21	0.51 - 0.54
Cherrington	GW *	58.55 – 58.83	0.28	NA	NA	0.76 – 0.82	0.88 – 1.10	NA
Edgmond	GW *	67.68 – 67.87	0.19	NA	NA	0.76 – 0.86	1.35 – 1.52	NA
Warren Farm	GW *+	77.02 – 77.38	0.36	NA	NA	0.61 – 0.67	1.58 – 1.92	NA
Gnosall	GW -	80.82 – 81.33	0.51	NA	NA	0.43 – 0.49	0.92 – 1.30	NA
Hawgreen	GW ×	70.35 – 72.20	0.85	NA	NA	0.86 – 0.88	0.46 – 0.95	NA
Heathlanes	GW ×	65.36 – 65.82	0.46	NA	NA	0.64 – 0.73	2.30 – 2.73	NA
Longdon	GW ×+	54.01 – 54.14	0.13	NA	NA	0.81 – 0.87	0.33 – 0.45	NA

BK × – Bridgnorth Kinnerton

WW - – Wildmoor Wilmslow

KC * – Kidderminster Chester Pebble beds

D + –Sand/gravel drift deposits

In general, the automatic calibration results for the homogenous model at the basin outlet show good results. For river flows, the automatic calibration at the basin outlet is improved compared to the manually calibrated model, with most internal gauging stations also showing improvement even though they were not used specifically in the calibration. The groundwater level results are not as good as the manually calibrated model (in which they were included within the calibration process), yet still show an ability to simulate the correct patterns of level simulation. The poorer overall accuracy at simulating the general water level is supported with the mean summary score for groundwater independently calculated at 0.60 as shown in Table 5.21, compared to 0.66 for the manually calibrated model (Table 5.8).

Table 5.21. Summary score results for groundwater levels at seven boreholes for the automatic calibration at the basin outlet for the homogenous model

	Cherrington	Edgmond	Warren	Gnosall	Hawgreen	Heathlanes	Longdon
General level	0.35	0.25	0.21	0.22	0.38	0.18	0.40
Overall shape	0.40	0.40	0.23	0.13	0.34	0.28	0.40
Total	**0.75**	**0.65**	**0.44**	**0.35**	**0.72**	**0.46**	**0.80**
Mean score	**0.60**						

5.5.3. Automatic calibration using multi-criteria and multi-location approach

The same methodology outlined in Section 5.5.1 has been used for the second automatic calibration that uses the objective function of equally weighted RMSE of flow at four river flow gauging stations, and another objective function of equally weighted RMSE of groundwater levels at seven boreholes as shown in Figure 5.19. The calibration resulted in convergence criteria being met after 281(+1 initial) simulations of the model that used different parameter sets in each run.

The mean optimal values of each RMSE objective function were 1.2773 m^3s^{-1} for flow, and 0.7779 m for groundwater. For comparison to the manual calibration and automatic calibration at Walcot, the optimal RMSE value for flow at Walcot that was achieved in this second automatic calibration was 3.030 m^3s^{-1} (RMSEP = 0.411) compared to the first autocal at the basin outlet, 2.574 (RMSEP = 0.349), and manual calibration 2.870 (RMSEP = 0.390).

5.5.3.1. Selection of calibrated models and parameter variation

As previously discussed, Madsen, (2003) used a similar approach of automatic calibration with two objective functions for the Karup catchment, Denmark, using a MIKE SHE model. He used a pareto optimum approach of selecting 19 feasibly calibrated models from the generated parameter sets provided by the Autocal procedure. Madsens work used, one objective function that assessed the RMSE from a set of gauging stations, and the second assessed the RMSE from a set of groundwater level boreholes.

In a similar method to Madsen, it would be possible to select parameter sets/models along a Pareto front (shown as an example in Figure 5.24). However, this option does not consider the models that are close behind the pareto front (that are very statistically similar in model performance), notably in the region where the RMSE for the two objective functions are both low, as shown by the green models in Figure 5.24.

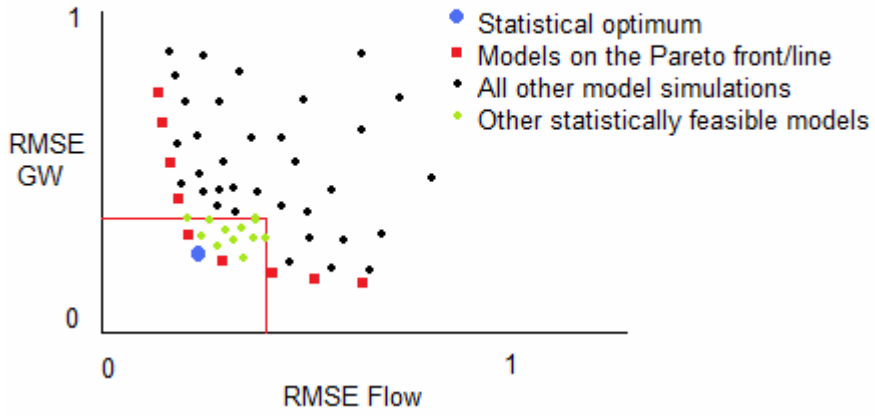


Figure 5.24. Example of the Pareto front

For this reason, a similar method of model selection to that described in the previous Section (5.5.2), is used where a given percentage from the optimal parameter sets RMSE is defined. When this method is used for this automatic calibration, this method allows parameter sets that have the lowest RMSE for both objective functions to be included as 'calibrated' models, and does not specifically highlight models/parameter sets that are better for simulating one or other of the objective functions. For example, the pareto front method may include calibrated parameter sets that result in good groundwater level calibration, but poorer river flow calibration.

In order to assess a realistic threshold for calibrated model selection Table 5.22 summarises the number of model RMSEs that are below different percentage thresholds. It can be seen that different numbers of models/parameter sets are found for the different river flow and groundwater level objective functions, with the number of models allotted within the river flow objective function greater than that of groundwater levels. The 5% threshold from optimal RMSE values for both flows and groundwater levels was selected. Although this threshold is higher than for the automatic calibration at the basin outlet in the previous section (2%), the higher threshold accommodates for the inclusion of two objective functions, not one. At the 2% level, only four models

satisfied the selection criteria. In total 28 models were selected that had RMSE values beneath the 5% thresholds for both river flows and groundwater levels. This equates to a total of 9.9% of all the models tested within the automatic calibration procedure.

To enable a comparison of the automatically calibrated models (in Chapter 7) the same method has been adopted as in the previous section; to report the results of a sampled set of ten of these models. Figure 5.25a shows the results of the selection process, with the optimal model, the 5% from optimal RMSE region (28 models) highlighted, and the further sample of ten of these models. Once again, the sample is systematically generated by listing the different models in run order, and selecting every fifth model from the list after ensuring the inclusion of the optimal model and the two lowest groundwater and flow RMSEs within the set of 28 models. The optimisation for both objective functions are shown in Figure 5.25b and c which plot the RMSE of the objective function against the simulation number to show how in both cases the RMSE has been optimised to the point in which a termination was met when the improvement in RMSE was not greater than 0.01 from a previous simulation. The plots also show the sampled suite of ten models within the context of all the simulations, as well as the 28 models within the 5% from optimal RMSE.

Table 5.22. Selection of 5% from optimal RMSE threshold of calibrated models

Multi-criteria/multi-location Homogenous setup – 282 parameter sets tested
Optimal mean River flow RMSE (m^3s^{-1}) = 1.2773
Optimal mean Groundwater level RMSE (m) = 0.7779

% increase from optimal		1%	2%	3%	4%	5%	10%
Increase allowed from optimal RMSE	Flow	0.0128	0.0255	0.0383	0.0511	**0.0639**	0.1277
	GWL	0.0078	0.0156	0.0233	0.0311	**0.0389**	0.0778
Threshold	Flow	1.2900	1.3028	1.1356	1.3283	**1.3411**	1.4050
	GWL	0.7856	0.7934	0.8012	0.8090	**0.8168**	0.8556
No. parameter sets within flow threshold		3	4	73	79	**90**	135
No. parameter sets within GWL threshold		3	4	33	41	**50**	95
No. parameter sets within both flow and GWL threshold		**3**	**4**	**8**	**17**	**28**	**76**
% of total runs		**1.1**	**1.4**	**2.8**	**6.02**	**9.9**	**27**

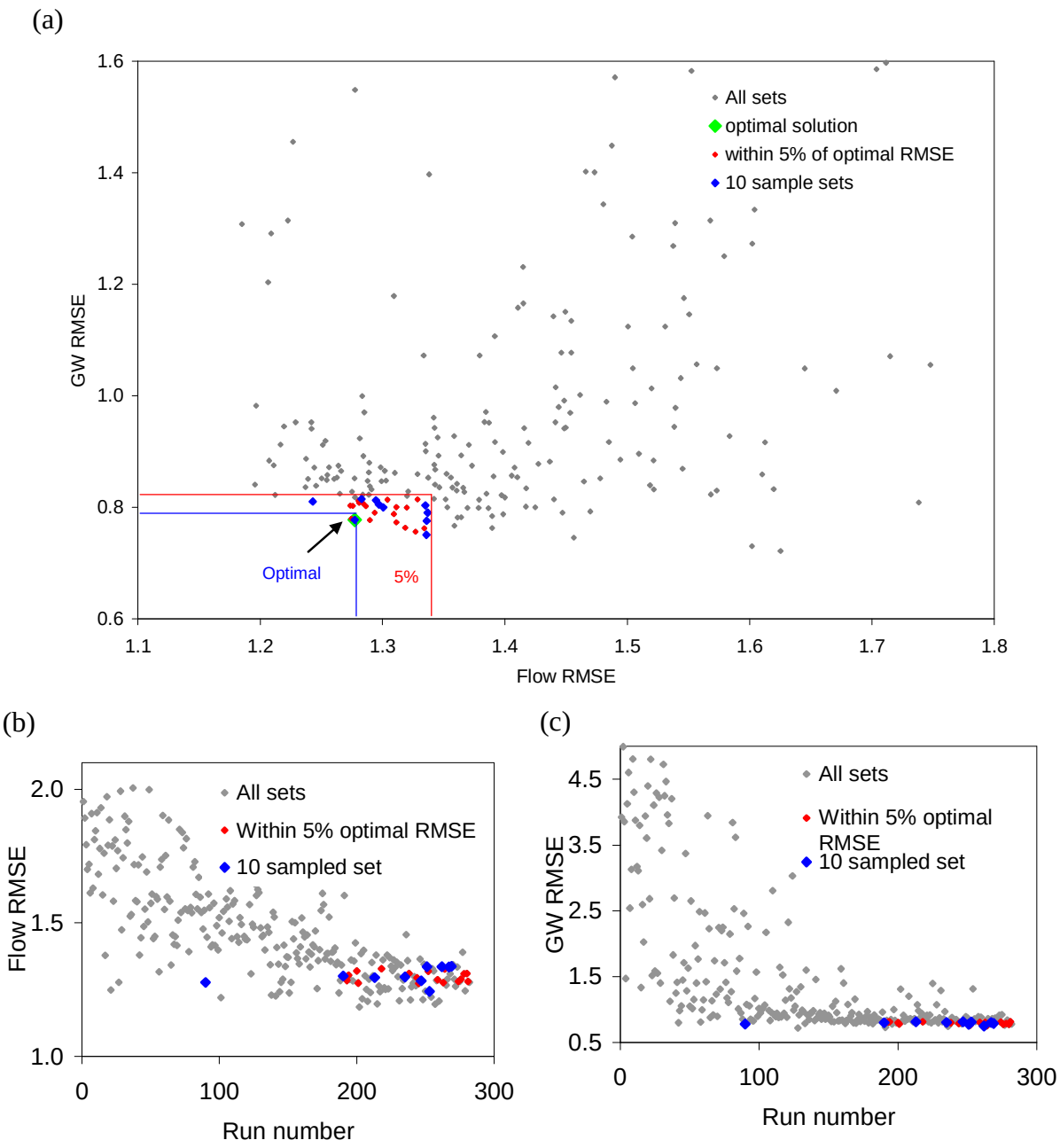


Figure 5.25. (a) Mean river flow and groundwater level RMSE identifying the optimal and sampled 'calibrated models' (b) Optimised mean river flow RMSE (m^3s^{-1}) (c) Optimised mean groundwater level RMSE (m)

Figure 5.26 presents the optimisation of the individual parameters that were subject to automatic calibration, with the selection of the suite of ten models, as well as all 28 models within 5% of the optimal RMSE also shown. Table 5.23 further summarises the calibrated parameter values for the sampled ten models which are also shown within the context of all 28 models for a comparison. It can be seen that the SWC Sat parameter shows the most refined optimisation near the upper limit of the defined parameter space.

The other parameters show less defined optimisation and as a result the optimal range of the parameters (from both the 28 calibrated and ten sampled models) are relatively wide compared to the automatic calibration just at the basin outlet. These wider parameter spaces may be due to these six parameters having to represent the entire heterogeneity of the catchment whilst trying to optimise based on a wide criteria (flows on different tributaries and groundwater levels from different geologies). As highlighted in Table 5.23, the range of each parameter value sampled in the ten models does not cover the whole limit as calculated by the 28 models within the 5% limit. However, the ten models do cover the calibrated ranges of flows and groundwater level RMSE variation of all 28 models.

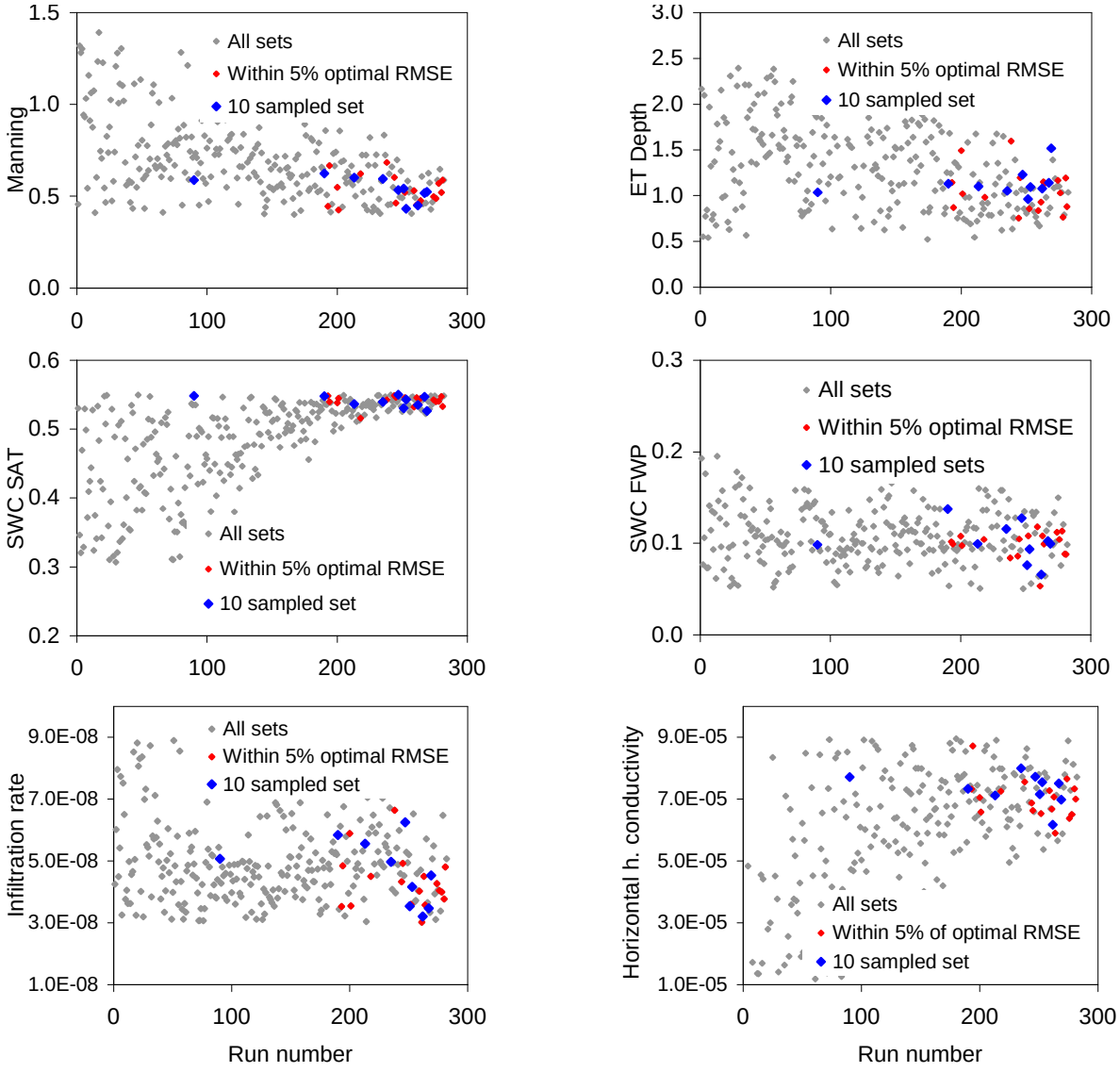


Figure 5.26. Parameter ranges tested within the automatic calibration using the multi-criteria approach and selected sample set of ten models for the homogenous model. Units for y axes: SWC Sat and SWC Fwp (fractions), infiltration rate and horizontal hydraulic conductivity (m s^{-1}), ET Depth (m)

Table 5.23. Parameter values for ten sampled models within the context of minimum and maximum parameter ranges for the multi-objective automatically calibrated homogenous model. Units: SWC Sat and SWC Fwp (fractions), infiltration rate and horizontal hydraulic conductivity (m s^{-1}), ET Depth (m)

Run	Manning surface roughness	SWC SAT	SWC FWP	Infiltration rate	Depth of ET	Horizontal hydraulic conductivity	Mean Flow RMSE (m^3s^{-1})	Mean GW RMSE (m)
90	0.587	0.548	0.098	5.072E-08	1.037	7.712E-05	1.2773	0.7779
190	0.623	0.548	0.137	5.839E-08	1.131	7.337E-05	1.3005	0.8000
213	0.600	0.536	0.099	5.556E-08	1.104	7.121E-05	1.2945	0.8129
235	0.595	0.540	0.116	4.974E-08	1.052	7.995E-05	1.2970	0.8049
247	0.534	0.550	0.128	6.255E-08	1.231	7.727E-05	1.2828	0.8151
251	0.541	0.530	0.076	3.536E-08	0.963	7.160E-05	1.3360	0.7757
253	0.431	0.543	0.094	4.164E-08	1.097	7.554E-05	1.2430	0.8106
262	0.449	0.535	0.066	3.202E-08	1.081	6.178E-05	1.3358	0.7506
267	0.518	0.547	0.103	3.472E-08	1.142	7.509E-05	1.3350	0.8037
269	0.524	0.526	0.100	4.533E-08	1.519	6.980E-05	1.3368	0.7906
Parameter ranges of the 10 sample sets (above) within 5% from RMSE								
Minimum	0.431	0.526	0.066	3.202E-08	0.963	6.178E-05	1.2430	0.7506
Maximum	0.623	0.550	0.137	6.255E-08	1.519	7.995E-05	1.3368	0.8151
Parameter ranges of all 28 calibrated sets within 5% from RMSE								
Minimum	0.425	0.516	0.053	3.025E-08	0.755	5.896E-05	1.2430	0.7506
Maximum	0.684	0.550	0.137	6.643E-08	1.596	8.717E-05	1.3368	0.8151

Having described the selection of calibrated models from the automatic calibration procedure to demonstrate equifinality, the output time series of river flows and groundwater levels and performance statistics from the ten sampled models are presented and discussed.

5.5.3.2. Calibrated model results and testing/validation

Figure 5.27 presents the calibrated bounds (upper and lower simulated MDF from the ten sampled models) for river flows at the four gauging stations used in the automatic calibration. Additionally, the observed river flow data and manually calibrated results are also shown to enable comparisons to be made. The calibrated bounds are wider than for the automatic calibration assessed using only the basin outlet (a mean range of 0.32 m^3s^{-1} at Walcot for the automatic calibration only at the basin outlet, increasing to a mean range of 0.62 m^3s^{-1} for this automatic calibration, Tables 5.20 and 5.26) with the largest range in output river flows seen during the winter months and during peak flow events. This widening in calibrated range is likely due to the inclusion of two different objective functions resulting in larger individual parameter ranges that did not optimise as effectively as in the autocalibration at the basin outlet.

The results for river flow in general show a reasonable fit with the observed data, with a slightly reduced ability when compared to the manual calibration in Figure 5.27. The results, as can be expected, do not appear as good as for the automatic calibration using only flow at Walcot or as good as the manual calibration, as supported by the statistics in Table 5.26. This is likely due to the addition of the groundwater objective function, thus the model seeks to find a balanced statistical optimum between the two criteria. As previously noted, the optimal individual RMSE for flow at Walcot is 3.03 m^3s^{-1}, with the RMSE range for the ten models being 3.03 to 3.35 m^3s^{-1} (shown in Table 5.26). This result is compared to the 2.574 result in the previous automatic calibration optimising only on river flow at the basin outlet. The summary score that was calculated separately from the other statistics is shown in Table 5.24 where it is supported that this multi-objective automatic calibration results in a slightly decreased ability of the model at simulating river flows. The mean score is 0.6725 compared to 0.695 for the manual calibration, and 0.7725 for the model automatically calibrated to flows at the basin outlet.

The simulated river flows at Bailey Brook (despite being included as an output measure in the flow objective function) again appear to under simulate baseflow, a feature apparent in all calibrations of the homogenous model. The statistics in Table 5.26 confirm this weakness with Nash-Sutcliffe NSE ranges from 0.29 – 0.37 for the suite of ten models. Nash-Sutcliffe results for the Meese tributary (up to this point always the best simulated) are lower for this calibration method, with ranges between 0.49 – 0.69. Again this is perhaps due to the addition of groundwater levels being included within the calibration method. It is interesting to note than even though the Meese tributary was not included in the previous automatic calibration using just the basin outlet, results were better than they are for this method where it is included as an output measure.

Table 5.24. Summary score for four river flow gauging stations in the multi-objective automatic calibration of the homogenous model

Flow	Walcot	Roden	Meese	Bailey Brook
Baseflow	0.20	0.24	0.150	0.10
Peaks	0.10	0.18	0.220	0.10
Timing	0.23	0.22	0.200	0.20
MDF	0.19	0.12	0.195	0.045
Total	**0.72**	**0.76**	**0.765**	**0.445**
Mean score	**0.6725**			

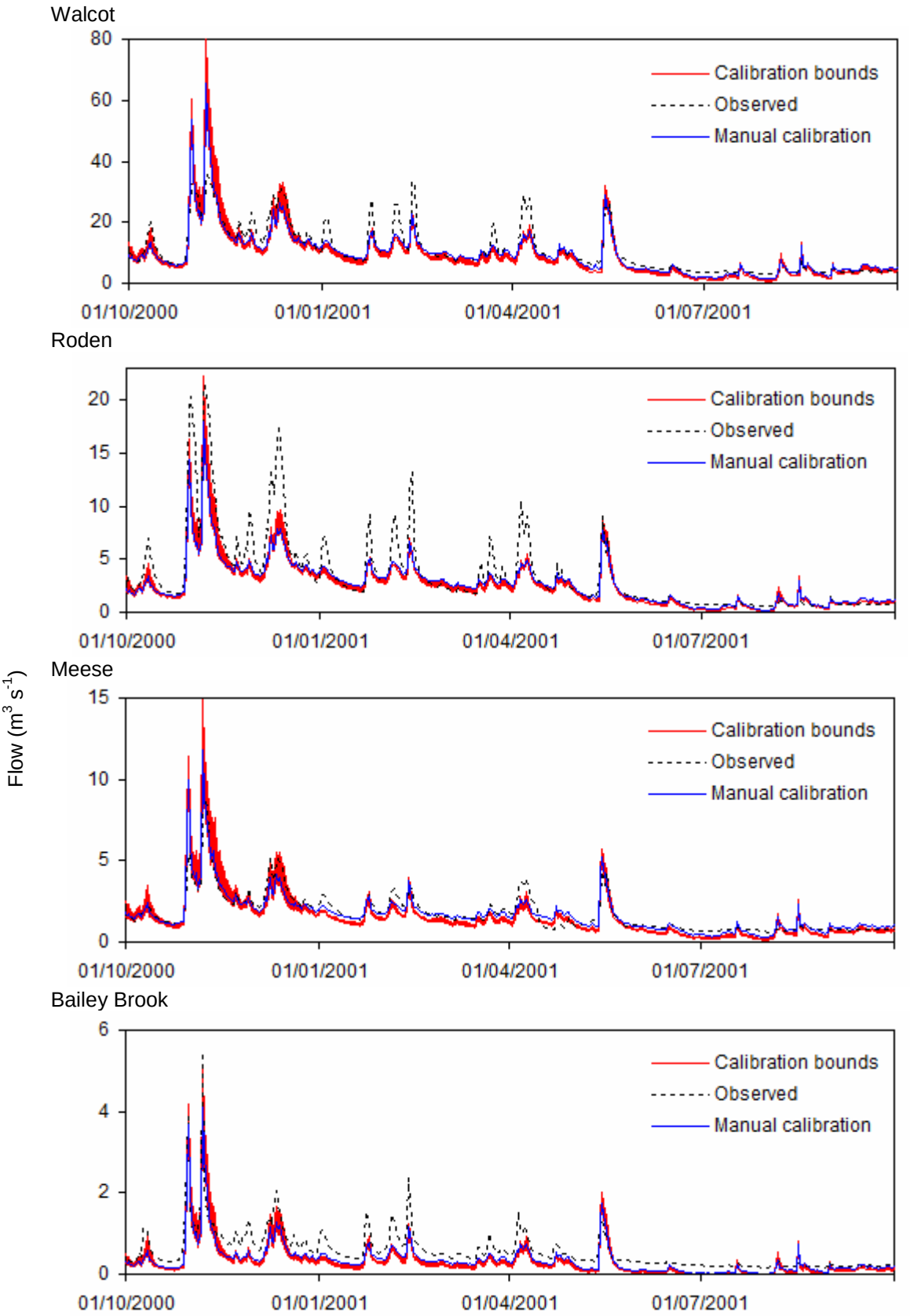


Figure 5.27. Observed and simulated daily river flow at four calibrated gauging stations. Calibration bounds (minimum and maximum daily flow) are shown for the ten sampled models, automatically calibrated with the multi-objective approach for the homogenous model

Figure 5.28 presents calibrated bounds of the ten sampled models for the seven groundwater level boreholes used as output measures in the groundwater objective function within this calibration. The observed data and manual calibration are also shown as a means of comparison. In all sites except Gnosall (that has been previously discussed as problematic to calibrate in the other methods), the models appears to show skill in reproducing the seasonal cycle and annual fluctuations, with an improvement from the manual calibration seen in all cases apart from Hawgreen. In addition, the general simulated levels are good, in most cases being within 2.0m of the observed levels.

The minimum and maximum range in calibration bounds are quite varied between boreholes, as highlighted in Table 5.26. The smallest mean simulated range at Edgmond is 0.35m compared to 1.04m at Gnosall. These values are compared to the same boreholes for the automatic calibration only using the basin outlet where the mean ranges of the ten models (max – min) are 0.19m and 0.51m respectively (Table 5.20). Despite this, Table 5.26 also indicates that the correlation, R, model performance at all groundwater boreholes is good, with the exception of Gnosall, with values typically between 0.65 to 0.91.

The calculation of the summary score for groundwater level boreholes is shown in Table 5.25. The mean score for all seven boreholes is 0.72, an improvement on the 0.60 score for the automatic calibration that used only flow at the basin outlet in the optimisation, and also an improvement on the 0.66 mean summary score for the manual calibration. The summary score therefore supports the result that groundwater levels are simulated best as a result of the multi-objective automatic calibration method. This is expected as groundwater RMSE has been included within the calibration and also indicates that when it is omitted from the calibration process the model performance of simulating groundwater levels is reduced.

Table 5.25. Summary score at seven groundwater level boreholes in the multi-objective automatic calibration of the homogenous model

	Cherrington	**Edgmond**	**Warren**	**Gnosall**	**Hawgreen**	**Heathlanes**	**Longdon**
General level	0.45	0.40	0.32	0.35	0.35	0.30	0.45
Overall shape	0.45	0.40	0.20	0.15	0.35	0.40	0.45
Total	**0.90**	**0.80**	**0.52**	**0.50**	**0.70**	**0.70**	**0.90**
Mean score	**0.72**						

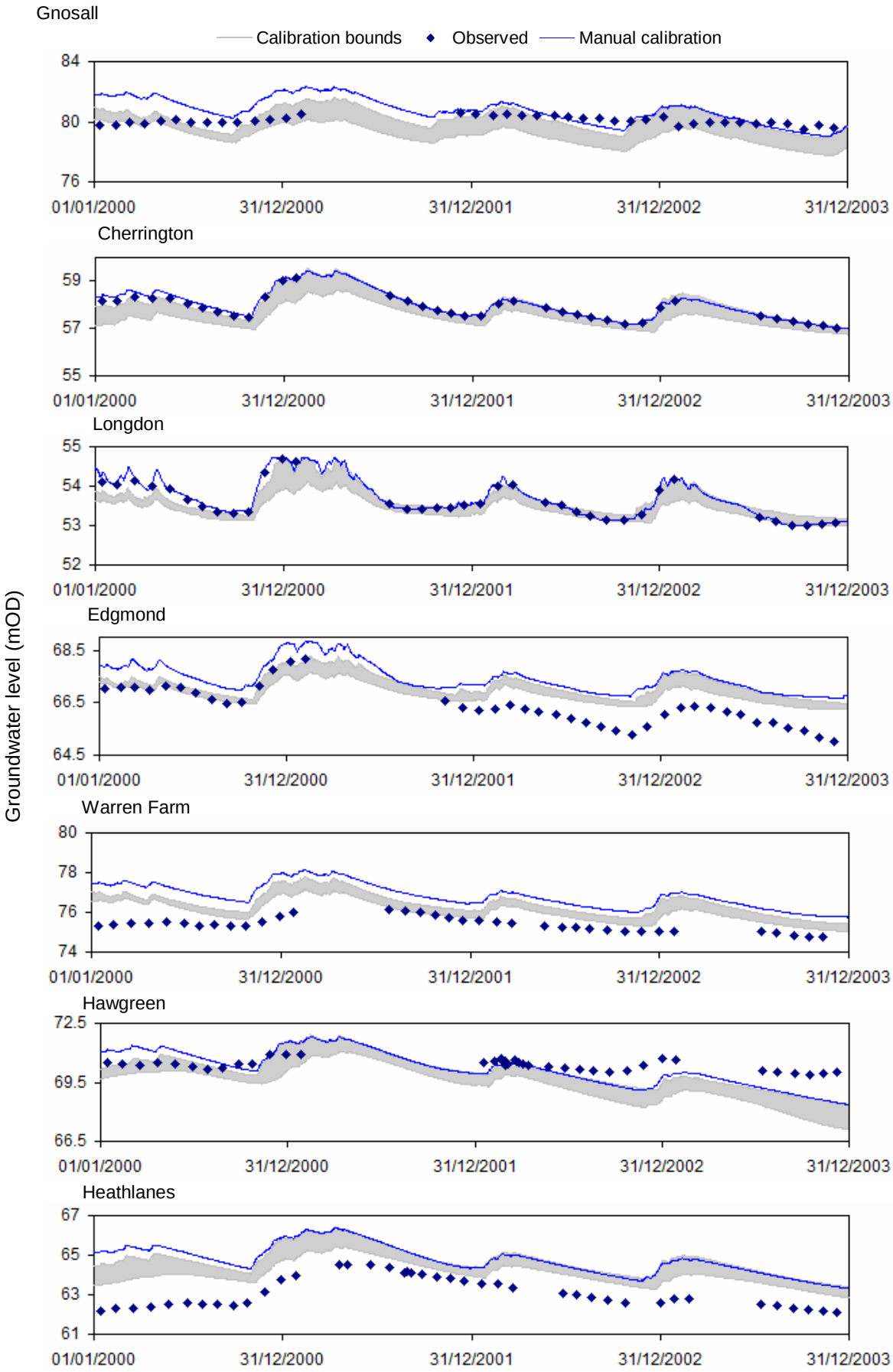


Figure 5.28. Results of the multi-objective automatic calibration of the homogenous model, manual calibration and observed levels from the range of ten sampled models are shown within the calibration bounds

Table 5.26. Statistical performance results from the automatic calibration of the multi-objective homogenous Tern model, for river flows and groundwater levels used in the objective functions/calibration. Results show the upper and lower (max and min) bounds for the range of the ten sampled models

Name	Measure	Min/max MDF (m^3s^{-1}) or MGWL (m)	Range (m^3s^{-1})or (m)	Total flow m^3s^{-1}	Nash-Sutcliffe NSE	Correlation R	RMSE Flow (m^3s^{-1}) GW (m)	RMSEP
Walcot	Flow at basin outlet	5.96 - 6.58	0.62	$7.23x10^8$ – $7.98 x10^8$	0.65 - 0.71	0.85 - 0.87	3.03 - 3.35	0.41 - 0.45
Roden	Flow largest tributary	1.60 - 1.75	0.15	$2.02 x10^8$ – $2.22 x10^8$	0.70 - 0.78	0.90 - 0.92	1.12 - 1.32	0.52 - 0.62
Meese	Flow 2nd largest tributary	1.00 - 1.12	0.12	$1.27 x10^8$ – $1.41 x10^8$	0.46 - 0.69	0.86 - 0.89	0.46 - 0.61	0.38 - 0.50
Bailey Brook	Flow at upstream tributary	0.19 - 0.21	0.02	$2.36 x10^7$ – $2.70 x10^7$	0.29 - 0.37	0.82 - 0.84	0.26 - 0.28	0.67 - 0.72
Cherrington	GW in KC	57.44 - 57.98	0.54	NA	NA	0.77 - 0.88	0.41 - 0.56	NA
Edgmond	GW in KC	66.82 - 67.17	0.35	NA	NA	0.74 - 0.84	0.64 - 0.91	NA
Warren Farm	GW in KC with sand/gravel D	76.02 - 76.42	0.40	NA	NA	0.60 - 0.77	0.67 - 1.00	NA
Gnosall	GW in WW	79.15 - 80.19	1.04	NA	NA	0.17 - 0.45	0.60 - 1.21	NA
Hawgreen	GW in BK	69.45 – 69.95	0.50	NA	NA	0.66 - 0.80	0.68 - 1.20	NA
Heathlanes	GW in BK	64.03 - 64.71	0.68	NA	NA	0.77 - 0.91	0.97 - 1.65	NA
Longdon	GW in BK with sand/gravel D	53.40 - 53.68	0.28	NA	NA	0.78 - 0.85	0.26 - 0.41	NA

BK – Bridgnorth Kinnerton, WW – Wildmoor Wilmslow, KC – Kidderminster Chester Pebble beds, D – Drift deposits

In order to assess the robustness of the multi-objective automatic calibration results, and how good simulations are at other non-calibrated sites, a testing/validation stage now uses the same additional gauging stations and groundwater boreholes that were used to test and validate the manual calibration in Section 5.3.

Results for the 'uncalibrated' river flows are shown in Figure 5.29 where a similar pattern to the manual calibration is seen. The flows at Eaton-on-Tern (with an NSE range of 0.67 – 0.75 and R between 0.87 – 0.89) and the flow at Ternhill (NSE ranging between 0.45 – 0.6, and R between 0.84 – 0.86) are comparable to the statistics shown in Table 5.10 for the testing and validation of the manually calibrated model in Section 5.3, (with NSE of 0.79 and 0.7. respectively, and R of 0.9 and 0.88).

Figure 5.29 also indicates that flow simulation at Potford Brook is not considerably improved as a result of this automatic calibration, with poor NSE statistics for the suite of ten models from -0.28 to -0.11. The model performance at Potford Brook was also poor for the manually calibrated model (in Section 5.3), and highlights that performance at uncalibrated gauging stations is not improved with the automatic calibration methodology.

Figure 5.30 compares the simulated groundwater levels to the observed data at the seven boreholes that are used for model testing (uncalibrated sites). The manually calibrated groundwater levels are also plotted for a comparison (from Figure 5.8) to assess if the automatic calibration procedure resulted in improved groundwater levels. It can be seen that the uncalibrated groundwater levels are improved at all boreholes except Woodlands Farm, with the automatically-calibrated suite of ten models showing closer fits with the observed data. A substantial improvement is seen in the simulated levels when compared to the manual calibration at Heathcote, with the RMSE ranging between 0.42 – 0.69 m (Table 5.27) compared to the higher RMSE from the manual calibration, 0.86 m (Table 5.11). This highlights that this calibration method is more robust than the manual calibration, especially with regard to the ability of simulating groundwater levels.

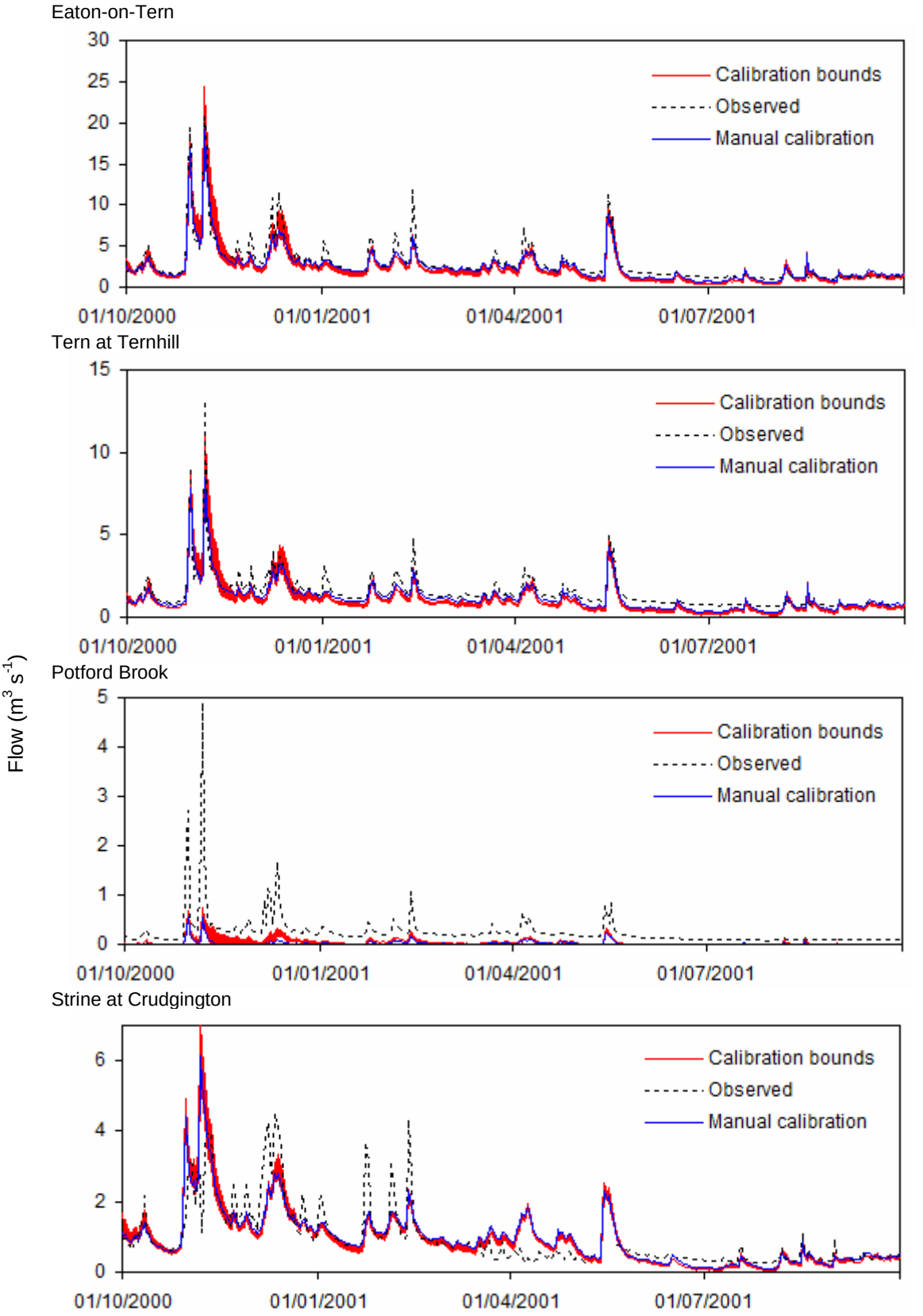


Figure 5.29. Validation/testing of the multi-objective automatic calibration of the homogenous model. Plots for four 'non-calibrated' gauging stations show the observed and simulated river flows (calibrated bounds derived from the min and max daily flow from the ten sampled models)

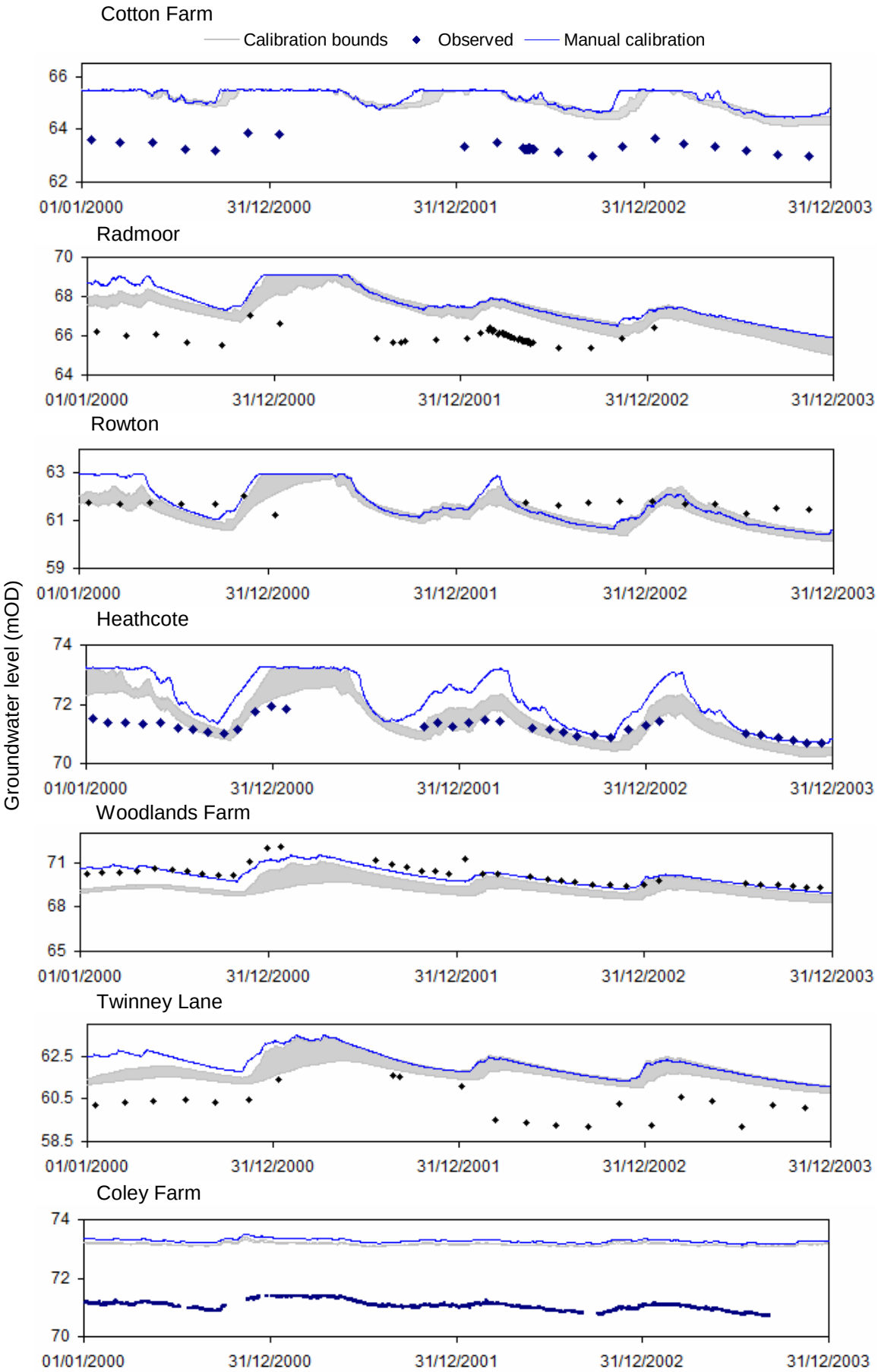


Figure 5.30. Testing/validation of the multi-objective automatic calibration for the homogenous model. Plots show observed and simulated levels at seven 'non-calibrated' boreholes (calibrated bounds derived from the simulated min and max levels from the sampled ten calibrated models)

Table 5.27. Statistical performance results from the testing/validation of the multi-objective homogenous Tern model, for river flows and groundwater levels not assessed within the calibration methodology. Results show the upper and lower (max and min) bounds for the range of the ten sampled models

Name	Measure	Min/max MDF (m^3s^{-1}) or MGWL (m)	Range (m^3s^{-1}) or (m)	Total flow m^3s^{-1}	Nash-Sutcliffe NSE	Correlation R	RMSE Flow (m^3s^{-1}) GW (m)	RMSEP
Eaton on Tern	Flow	1.57 - 1.77	0.20	1.98×10^8 – 2.24×10^8	0.67 - 0.75	0.87 - 0.89	0.81 - 0.94	0.42 - 0.48
Tern at Ternhill	Flow	0.73 - 0.85	0.12	9.24×10^7 – 1.07×10^8	0.45 - 0.60	0.84 - 0.86	0.44 - 0.51	0.42 - 0.49
Potford Brook	Flow	0.01 - 0.02	0.01	1.30×10^6 – 2.59×10^6	-0.28 -0.11	0.71 - 0.75	0.22 - 0.24	1.38 - 1.50
Strine	Flow	0.60 - 0.65	0.05	7.52×10^7 – 8.27×10^7	0.51 - 0.59	0.78 - 0.79	0.38 - 0.41	0.53 - 0.57
Cotton Farm	GW in BK with silty clay D	65.03 - 65.15	0.12m	NA	NA	0.78 - 0.83	1.69 - 1.82	NA
Radmoor	GW in BK with diamicton D	67.03 - 67.49	0.46m	NA	NA	0.46 - 0.63	1.35 – 1.79	NA
Rowton	GW in BK silty clay D	61.22 - 61.70	0.48m	NA	NA	-0.09 - 0.07	0.59 - 0.83	NA
Heathcote	GW in BK with sand/ gravel D	71.34 - 71.75	0.35m	NA	NA	0.84 - 0.90	0.42 - 0.69	NA
Woodlands Farm	GW in KC	69.02 - 69.65	0.63m	NA	NA	0.55 - 0.73	0.88 - 1.40	NA
Twinney Lane	GW in KC	61.50 - 62.05	0.55m	NA	NA	0.21 - 0.39	1.37 - 1.84	NA
Coley Farm	GW in WW with sand/ gravel D	73.15 - 73.17	0.02m	NA	NA	0.67 - 0.75	2.10 - 2.12	NA

BK – Bridgnorth Kinnerton, WW – Wildmoor Wilmslow, KC – Kidderminster Chester Pebble beds, D – Drift deposits

5.5.4. Review of the automatic calibration of the homogenous model

To summarise the results of the automatic calibrations by referring back to the objectives outlined at the beginning of this section, the primary aim was to test whether the automatic calibration methodology resulted in improved simulations of river flows and groundwater levels. The calibrated minimum and maximum bounds for each set of ten models have been directly compared to the manually calibrated model, where it has been shown that the automatic calibration methods – both at the basin outlet and the multi-objective method have produced slightly improved quantitative results than the manual calibration. Although the manually calibrated results are very comparable, the autocal methodology, as expected, results in a statistically balanced calibration according to the objective function, where as the manual calibration is a subjective 'best fit'.

The secondary objective was to compare the complexity of the automatic calibration methods, and whether by including more rigorous objective functions (in the second method) the ability of the model simulations of flow and groundwater are improved. Results have shown that there is an apparent trade off in simulation ability with the two different automatic calibrations. As could have been expected, the automatic calibration at the basin outlet resulted in improved simulations of river flow at the basin outlet compared to the manual calibration and the multi objective automatic calibration. The RMSE value at the basin outlet from the manually calibrated model was 2.87 m^3s^{-1}, improved to between 2.57 – 2.62 m^3s^{-1} for the ten automatically-calibrated models at the outlet and reduced to 3.03 – 3.35 m^3s^{-1} in the multi-objective automatic calibration. More generally for the other gauging stations, there were also slightly improved RMSE and Nash-Sutcliffe NSE values when automatically calibrated at the basin outlet, when compared to the other two methods of calibration.

However, the simulated groundwater levels that were effectively un-calibrated in the automatic calibration at the basin outlet resulted in slightly poorer simulations when compared to the manual calibration. Despite this, the general groundwater level performance was still comparable to the manual calibration. The second multi-objective and multi-site automatic calibration that included groundwater in an objective function resulted in the best simulations of groundwater of the three calibrations tested.

Given the statistical results as well as the results of qualitative assessment (expert elicitation) and derivation of the summary score, the results suggest that the different calibration methods can all be considered to perform well for different aspects of the model. The qualitative assessment indicates that the multi-objective automatic calibration results in the most balanced model as it was most robustly calibrated; however this is at a trade off of resulting in the output flow statistics being the poorest of the three methods tested. It is also relevant to note that the model performance at some sites (e.g. Potford Brook river flow, and groundwater levels at Gnosall) are not notably different for any of the calibration methodologies assessed. This highlights that the problems in performance at these sites may not be due to the calibration but perhaps due to the homogenous nature of this model; that the parameter set for the catchment are not representative of the processes at these sites. As already discussed, the problems may also be due to scaling (Potford Brook) and the close to boundary location (Gnosall). This will be assessed in the analysis of output results from the distributed model in the following chapter.

Lastly, the automatic calibrations sought to assess and quantify some of the parameter equifinality within hydrological modelling. This was undertaken using defined thresholds – that if a parameter set resulted in a model within a prescribed percentage of the optimal RMSE, then it was considered as a calibrated model. Ten models each with different parameter sets were sub-sampled, plotted and compared for each automatic calibration method.

In the presented results it has been shown that parameter equifinality is larger in the multi-objective automatic calibration, which includes two objective functions. The multi-objective automatic calibration yields wider calibrated output bands compared to the automatic calibration using only at the basin outlet. Additionally there is also more variation within the calibrated values within the parameter space, despite the mean flow and groundwater RMSE values differing from the optimum value only marginally.

5.6. Summary

This chapter focussed on the calibration and sensitivity analysis of the homogenous Tern model that was described in Chapter 4. The performance criteria by which the

models were assessed were described with the introduction of the summary score measure, followed by initial tests and manual calibration and validation of flows and groundwater levels. In order to better understand the model's processes and to derive key sensitive parameters in subsequent automatic calibration tests, a manually set up sensitivity analysis of model parameters was undertaken with seven sensitive parameters identified. Two automatic calibration methods were then undertaken within the MIKE ZERO Autocal software that compared the automatic calibration at the basin outlet with an automatic calibration using a multi-location and multi-criteria approach.

The three methods of calibration have all shown ability to reproduce flows and groundwater well with minimal differences. In all cases, the testing and validation stages show the model performance to be satisfactory but not as good as during calibration. In the following chapter, a similar structure and tests are adopted for the distributed Tern model.

Chapter 6

Initial calibration, sensitivity analysis & automatic calibration of the distributed Tern model

6.1. Introduction

This chapter reviews the setup, calibration and testing of the Tern Distributed Model that was detailed in Chapter 4. Initial setup and simulations, as well as the manual calibration and validation of river flows and groundwater levels are described in Section 6.2. The calibration and testing use the same performance criteria as the homogenous model, previously introduced in Section 5.2.

An automatic parameter sensitivity analysis is then described in Section 6.3 that has been undertaken within the Autocal component in MIKE ZERO. Section 6.4 follows the same structure as the previous chapter, detailing the automatic calibration of the model for both the basin outlet at Walcot, and a second automatic calibration using a number of river flow gauging stations and groundwater level boreholes.

6.2. Initial testing and manual calibration

As with the homogenous model, the purpose of a manual calibration was to obtain an adequately calibrated model to then be carried forward for parameter sensitivity analyses (Section 6.3). The completed set up and first simulation of the distributed Tern model initially used the manually calibrated parameter values from the homogenous model, but specified according to the distributed set-up of the model. For example, each of the five unsaturated zone classes in the distributed model were set with the same parameter values (e.g. Sat = 0.5, Fc = 0.4, Fwp = 0.18). This process allowed for a test of the new model construction and run errors before more specific distributed parameter values were specified to the model for each class.

The model parameter values were then spatially varied to be catchment specific, with values specified within the allowed parameter limits shown in Table 6.1. As reviewed in Chapter 4, these parameter ranges were derived from a range of methods including secondary data obtained from field research (e.g. of specific yield values Johnson, 1967 and LOCAR, 2000), from literature that was both site or region specific (such as horizontal and vertical hydraulic conductivity and specific yield in the saturated zone class derived from BGS reports) as well as wider literature that sought to classify typical values of LAI and rooting depths of specific land covers (e.g. Canadell et al., 1996). In cases where specific values could not be attained for parameters (such as the storage coefficient in the saturated zone), then the parameter was represented by a best estimate and were subject to later sensitivity analyses.

The process of manual model calibration and adjustment of each parameter value was iterative (as with the homogenous model), where parameters were perturbed gradually to consequently understand the response on model outputs. Specific attention was given to the saturated zone parameters that had increased from one uniform class to ten classes (four drift classes and six solid geology classes).

Again using the multi-site and multi-proxy method of assessment (Refsgaard, 1997; 2000), the model outputs of river flow and groundwater levels at different sites were assessed after each adjusted simulation, with the aim of achieving best possible RMSE, NSE and R statistics.

Table 6.1. Review of the distributed model set-up and parameterisation resulting for the manual calibration. ** denotes where a parameter is also subject to sensitivity analysis in Section 6.3

MIKE SHE Component	**Parameter**	**Representation**		
Model Domain	Grid size	1000m		
Topography	Topography	1000m grid		
Precipitation	Uniform according to daily Tern catchment areal mean	Time varying, spatially distributed at 8 sites by Thessien Polygons		
Land use (uniform)	LAI Root depth	Spatially varying according to 8 land use classes (with agricultural and broadleaf forest classes time varying according to the growing season) Table 4.6		
Evapo - transpiration	Uniform according to calculated Hargreaves-Samani data	Time varying but uniform value used spatially within the catchment		
Rivers & Lakes	MIKE 11	Constant (described in Chapter 4)		
		Lower limit	**Upper limit**	**Calibrated parameter value**
Overland flow	Manning number	0.1	1.5	1) 0.42 2) 0.42 3) 0.38 4) 0.32 5) 0.30 6) 0.30 7) 0.15 8) 0.20
	Detention storage (mm)			** 0.15 mm
	Initial water depth (mm)			** 0.005 mm
UZ flow	Soil water content at saturated conditions	0.345	Defined for 5 classes ** 1) 0.386 – 0.549 2) 0.345 – 0.526 3) 0.378 – 0.546 4) 0.354 – 0.592 5) 0.372 – 0.535	 0.455 0.434 0.442 0.460 0.434
	Soil water content at field capacity	0.098	Defined for 5 classes ** 1) 0.290 – 0.456 2) 0.265 – 0.425 3) 0.145 – 0.370 4) 0.098 – 0.531 5) 0.105 – 0.394	 0.380 0.339 0.277 0.325 0.255
	Soil water content at field wilting point	0.026	Defined for 5 classes ** 1) 0.140 – 0.309 2) 0.117 – 0.269 3) 0.034 – 0.154 4) 0.026 – 0.322 5) 0.028 – 0.171	 0.231 0.187 0.101 0.162 0.090
	Infiltration rate	4.05e-8	Defined for 5 classes ** 2) 4.51e-8 – 6.37e-8 3) 4.05e-8 – 6.13e-8 4) 4.50e-8 – 6.37e-8 5) 4.05e-8 – 6.83e-8 6) 4.28e-8 – 6.25e-8	 5.3e-8 5.0e-8 5.1e-8 5.3e-8 5.0e-8
	Depth of evapotranspiration (m)	0.5	Uniform ** 0.5 – 2.5	 ** 2 m

SZ flow	Lower level (relative to ground surface level)			-100m surface elevation
	Horizontal hyd raulic conductivity ($m\ s^{-1}$)	1.0e-8	Defined for 10 classes **	
			D1) 1.0e-5 – 9.0e-5	5e-5
			D2) 1.0e-5 – 9.0e-5	5e-5
			D3) 1.0e-4 – 5.7e-4	0.00028
			D4) 1.0e-8 – 1.0e-7	5.78e-8
			S1) 1.0e-8 – 5.0e-6	1e-6
			S2) 5.0e-5 – 9.0e-5	7e-5
			S3) 1.0e-5 – 9.72e-5	2.08e-5
			S4) 1.5e-5 – 9.72e-5	3e-5
			S5) 2.0e-5 – 1.736 e-4	6.82e-5
			S6) 1e-5 – 9.72e-5	2.08e-5
	Vertical hydraulic conductivity ($m\ s^{-1}$)	1.0e-11	Defined for 10 classes **	
			D1) 9.0e-9 – 5.0e-5	9e-6
			D2) 9.0e-9 – 5.0e-5	9e-6
			D3) 9.0e-9 – 5.0e-5	9e-6
			D4) 9.0e-9 – 5.0e-5	9e-6
			S1) 1.0e-11 – 5.0e-8	2e-8
			S2) 1.0e-9 – 5.0e-7	2e-8
			S3) 4.86e-8 – 1.04e-5	1e-7
			S4) 4.86 e-8 – 2.0e-5	8.5e-6
			S5) 5.68e-8 – 2.26e-5	2e-7
			S6) 5.78e-8 – 1.04e-5	1e-7
	Specific yield	0.1	Defined for 10 classes **	
			D1) 0.1 – 0.3	0.16
			D2) 0.1 – 0.6	0.44
			D3) 0.1 – 0.3	0.25
			D4) 0.1 - 0.3	0.12
			S1) 0.1 - 0.35	0.3
			S2) 0.1 - 0.35	0.25
			S3) 0.1 - 0.35	0.242
			S4) 0.1 - 0.35	0.232
			S5) 0.1 - 0.35	0.232
			S6) 0.1 - 0.35	0.265
	Storage coefficient ($m\ s^{-1}$)	5.0e-5	Defined for 10 classes **	
			D1) 5.0e-5 – 0.25	1e-4
			D2) 5.0e -5 – 0.25	1e-4
			D3) 5.0e -5 – 0.25	1e-4
			D4) 5.0e -5 – 0.25	1e-4
			S1) 5.0e -5 – 0.3	0.001
			S2) 5.0e -5 – 0.3	0.001
			S3) 5.0e -5 – 0.3	0.001
			S4) 5.0e -5 – 0.3	0.001
			S5) 5.0e -5 – 0.3	0.001
			S6) 5.0e -5 – 0.3	0.001
	Initial Potential head (m)			-2m surface elevation

Where: D1= Till/diamicton, D2= peat, D3= glacio-fluvial sands and gravels, D4= alluvial silty clay, S1= low permeability mudstones, S2= mixed permeability class, S3= Bromsgrove/Helsby Formation, S4= Kidderminster/Chester Pebble Beds, S5= Bridgnorth/Kinnerton Formation, S6= Wildmoor/Wilmslow Formation.

Specific attention was given to outputs at a particular gauging stations or groundwater level borehole if it was known to be located within the area for the parameter being adjusted. For example, river flow at Bailey Brook and in the Roden were given more attention when varying the S1 (low permeability mudstones) class, as the upper parts of these tributaries are located within the spatial extent of this class, and therefore it is within these tributaries that the model would be sensitive to parameter variation.

Despite the knowledge that had been acquired from model conceptualisation, and the sensitivity analysis and experience of calibrating the homogenous model, with so many permutations and possible parameter adjustments possible in the distributed model, it was difficult to find a parameter set which resulted in a good balance within the allowed parameter space, whilst resulting in acceptable model outputs. However, the selected parameter values shown in Table 6.1 were found to result in a model that was adequately balanced and carried forward for a parameter sensitivity analysis. Sections 6.2.1 to 6.2.4 describe the results of this best manual calibration, for both river flows and groundwater levels.

6.2.1. River flow assessment

Section 5.2 outlined the performance criteria by which the different calibrations of homogenous models were assessed. The distributed models in this chapter are also assessed using the same criteria to ensure the same comparison of model performance. Figure 6.1 presents observed and simulated daily river flows, and Figure 6.2 mean monthly river flows at the gauging stations used in model calibration. Inspection of the daily flow simulations suggests a reasonable fit with the observed data for all gauging stations.

The general peak river flows are simulated in a similar way to the manual calibration of the homogenous model (Figures 5.1 and 5.2), with the peak flows over-estimated at Walcot in the winter of 2000-2001 which was characteristically wet, and under simulated in 2002-3 which was characteristically dry with a much lower total annual precipitation. The mean monthly plot for Walcot at the basin outlet also summarises this pattern.

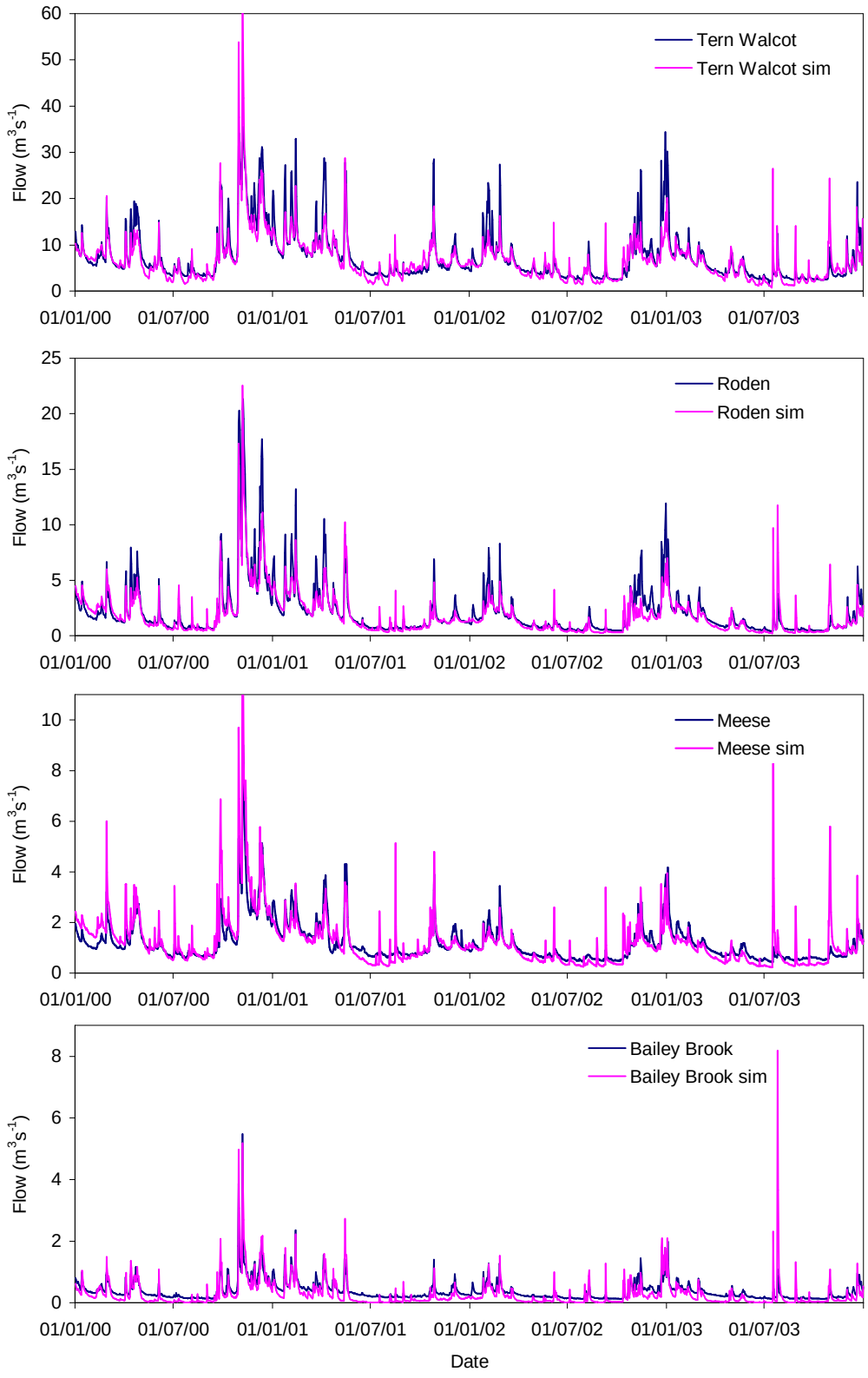


Figure 6.1. Observed and simulated daily river flows at four gauging stations used in the manual calibration of the distributed model

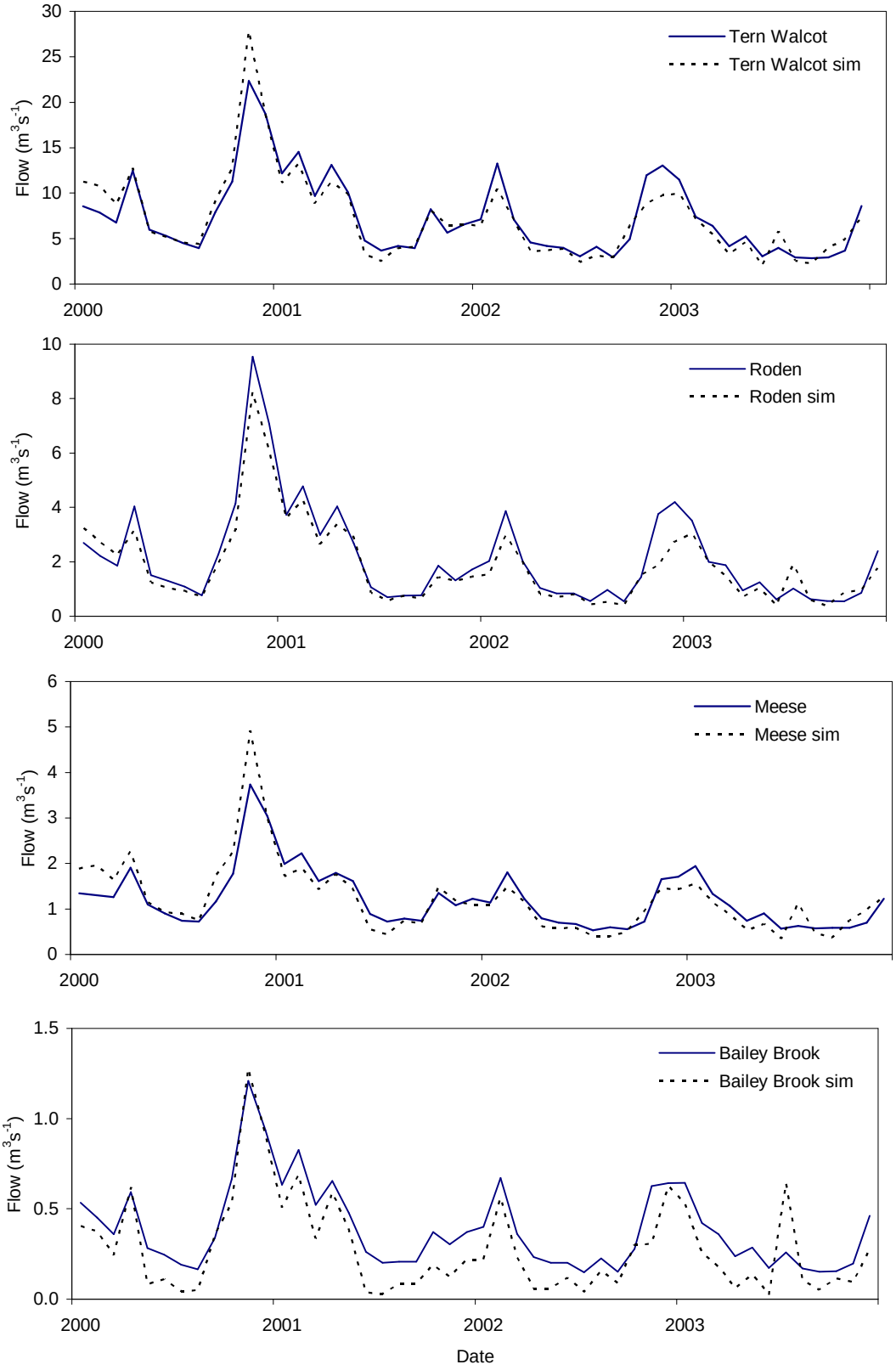


Figure 6.2. Observed and simulated mean monthly river flows for the four gauging stations used in the manual model calibration of the distributed model

There is a notable improvement of the peak flow simulation at Bailey Brook when compared to the homogenous model manual calibration (Figure 5.1). These improvements may be due to a better representation of input precipitation (from the Sandford raingauge as compared to the catchment average) but also due to a better representation of the unsaturated and saturated zone parameter values being more specific to the characteristic less permeable nature of this tributary compared to the otherwise sandstone dominated catchment. The under simulation of baseflows at Bailey Brook are again simulated in the distributed model, perhaps confirming the under simulations are due to scaling effects and the inadequate use of a 1000m grid cell representation in this small tributary (or also water balance error that was discussed in Section 3.10).

The summer baseflow component of the simulations appears good at Walcot and is best represented in the Roden tributary. However, the model under simulates river flow in Meese and Bailey Brook tributaries with the baseflow being lower than the observed. When reviewing the mean monthly plots in Figure 6.2 for the Meese, this is however less apparent and has likely been 'averaged out' (therefore highlighting the benefit of additionally showing the daily output plots of simulated flow at different sites). Although the general baseflow pattern is under simulated in the Meese, there are simulations of small peaks during summer months throughout the period that were not observed (nor were they simulated in the manual and automatic calibrations of the homogenous model).

It is likely that the over-simulated summer-time river flow peaks result from the inclusion of the eight individual rainfall timeseries from different sites that affect the volumes simulated in the nearby tributaries. For example, it is apparent there are very large over simulations of river flow in July 2003 (Figure 6.1). The over simulation is most pronounced in Bailey Brook but also simulated in the other sites. As shown in Figure 6.3, the Sandford raingauge covers the area of Bailey Brook where a simulated peak flow in 25/7/2003 of $8.19 m^3 s^{-1}$ is compared to an observed peak of $0.8 m^3 s^{-1}$. Figure 6.4 compares daily rainfall at all gauges as well as the catchment mean, for July 2003 when these over simulations occur. On 25/7/2003, 70.3 mm rainfall is recorded to have fallen at Sandford. However, this amount is shown to be far greater than at any other site and in Section 3.6.1.2 this event was flagged as a suspect record. The large

simulated flow is therefore attributed to be a result of this suspect input rainfall. Figure 6.4 also shows the catchment mean rainfall and it is apparent that on the same date, the catchment mean input rainfall to the homogenous model was much lower, and probably explains why these large river flow events are not shown.

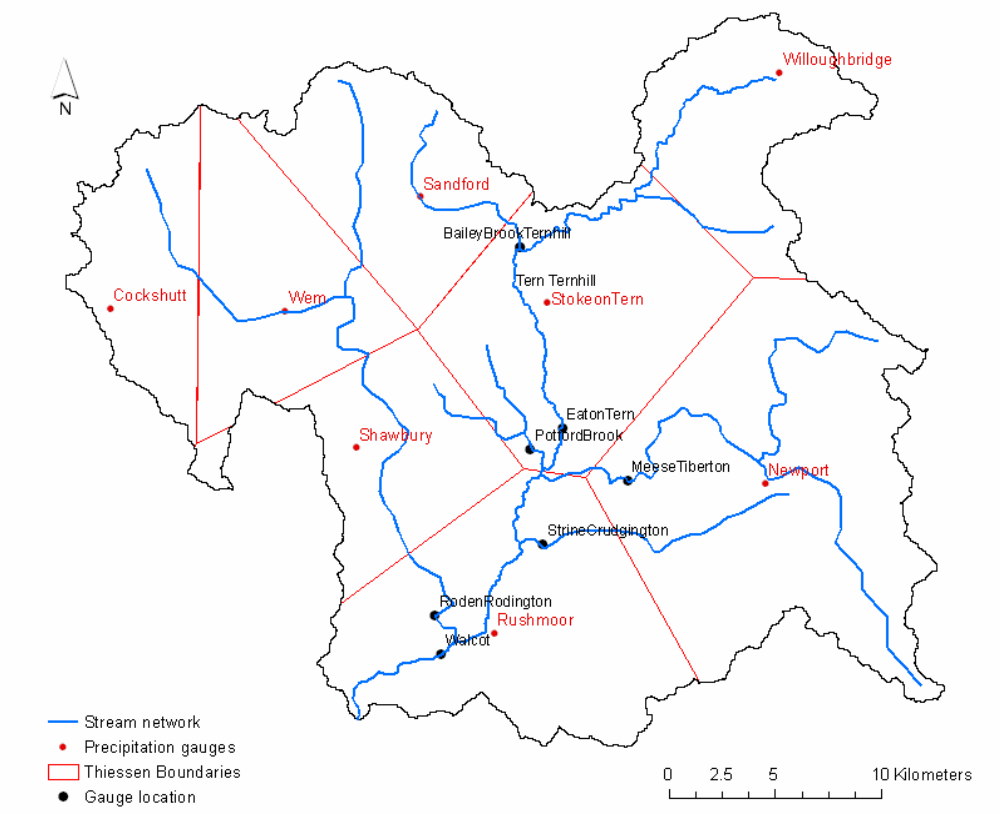


Figure 6.3. Thiessen polygon areas in the distributed model

The other tributaries, notably the Roden (covered by Cockshutt and Wem rain gauges), and the Meese (Newport raingauge), are also subject to very high river flows in July 2003 that are not observed in the daily observed flow records. Cockshutt recorded 70 mm of rainfall on 17th July 2003 (also flagged in Table 3.15 as suspect), a far greater amount than the other records, and Wem recorded higher than the catchment mean for both the 16th and 17th July 2003. These findings highlight the necessity of accurate rainfall data as inputs in hydrological models, and flags the possible outcome (resulting in high simulated peak flows) when potentially suspect or erroneous data are used.

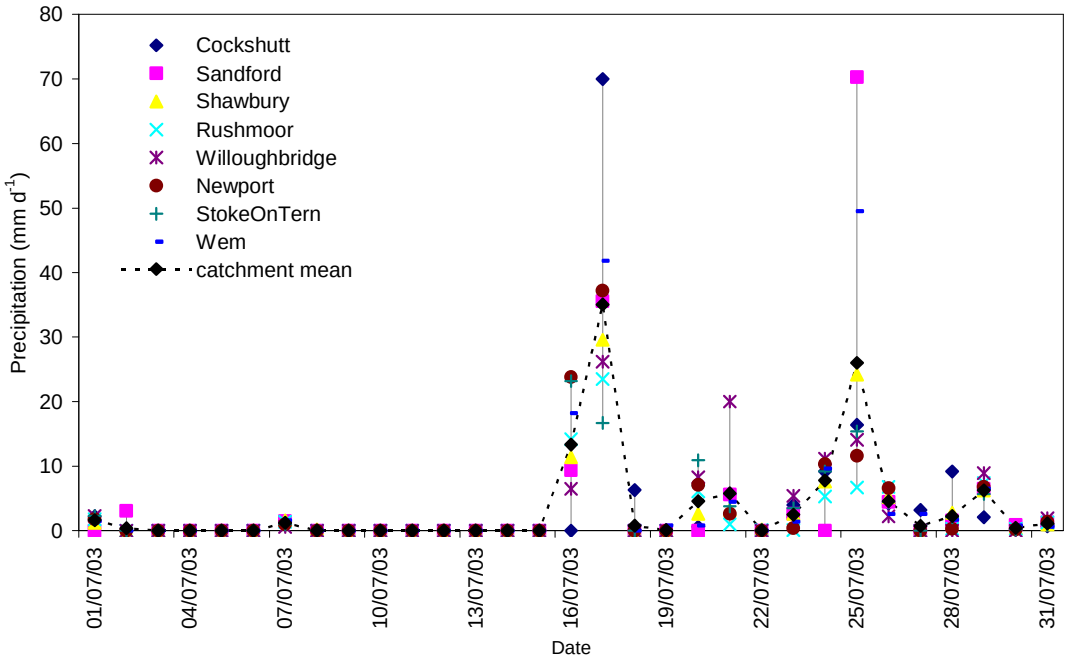


Figure 6.4. Comparative July 2003 daily precipitation record for the eight rainfall gauges used in the distributed model and the catchment mean as a comparison to the data used in the homogenous model

The calculated summary score results for the manual calibration of the distributed model that are given in Table 6.2 support the discussion that the overall performance of the calibrated model for river flows is good, with a mean score of 0.710 (improved from 0.695 in the homogenous model manual calibration). The individual performance at the basin outlet and in the Roden is 0.8, with the Meese and Bailey Brook scored relative to the problems seen in the peaks during summer months/baseflow (0.11 and 0.10 respectively) that bring down the overall score for each of these two tributaries. As with the homogenous model, the necessity to assess internal model performance is again shown, with the varying abilities of the model to simulate river flow in different tributaries compared to what would otherwise be a very good model just at the basin outlet.

Table 6.2. Results of summary score for manual river flow calibration of the distributed model

Flow	**Walcot**	**Roden**	**Meese**	**BB**
Baseflow	0.21	0.23	0.11	0.10
Peaks	0.11	0.16	0.10	0.20
Timing	0.24	0.23	0.20	0.20
MDF	0.24	0.18	0.24	0.09
Total	**0.80**	**0.80**	**0.65**	**0.59**

Catchment mean summary score = 0.71

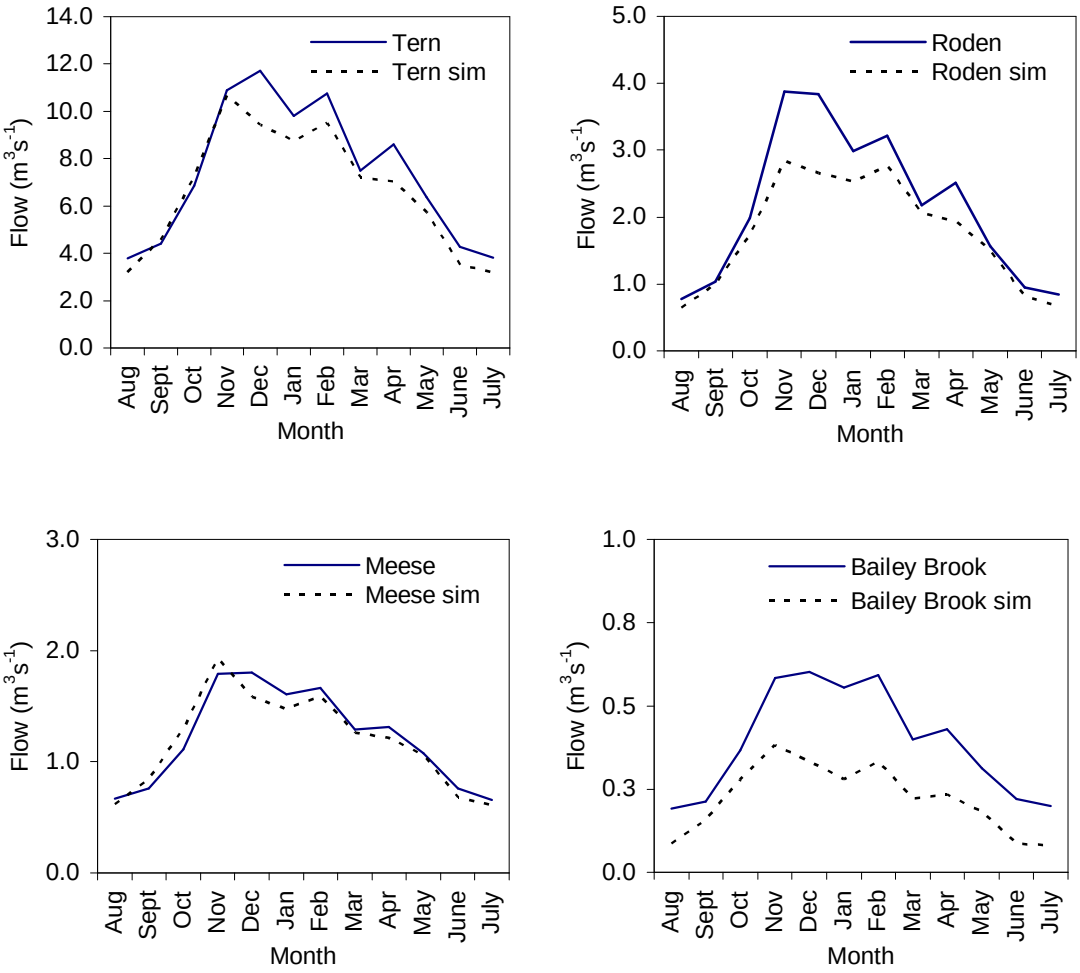


Figure 6.5. River flow regimes at the four gauging stations for the manual calibration of the distributed model

Table 6.3. Summary statistics of manually calibrated flow in the distributed model (Observed values in bold, simulated values in normal text)

Gauge	MDF m^3s^{-1}	Max flow m^3s^{-1}	Total flow volume m^3	R	NSE	RMSE m^3s^{-1}	RMSEP
Tern at Walcot	**7.37**	**35.44**	**9.31×10^{8}**	0.87	0.71	3.07	0.42
	7.26	78.08	9.16×10^{8}				
Roden at Rodington	**2.14**	**21.32**	**2.70×10^{8}**	0.91	0.81	1.05	0.49
	1.87	22.57	2.35×10^{8}				
Meese at Tibberton	**1.21**	**9.02**	**1.52×10^{8}**	0.85	0.5	0.58	0.48
	1.22	14.34	1.53×10^{8}				
Bailey Brook at Ternhill	**0.39**	**5.48**	**4.89×10^{7}**	0.80	0.13	0.31	0.79
	0.28	8.19	3.54×10^{7}				

Figure 6.5 summarises the river flow regimes at the four sites for the period 2000-2003. The regimes simulated at Walcot and in the Meese are good, although an under-simulation is seen during peak winter months, especially in the Roden, with under-simulation consistently shown at Bailey Brook.

On inspection of the calculated statistics in Table 6.3, the general skill of the model is confirmed, with comparative mean daily flows and total simulated flow compared to observed. Nash-Sutcliffe coefficients of 0.71 and 0.81 the simulations at Walcot and in the Roden, respectively, indicate good model performance. The Meese and Bailey Brook NSE are lower, and it is expected this is due to the suspect or erroneous input rainfall data in July distorting the statistic for the whole simulation. This result also confirms the necessity to plot the output timeseries of daily river flow, monthly mean and annual regimes as well as computing summary statistics, as the overall performance at Bailey Brook is improved compared to the manual calibration of the homogenous model, yet the NSE statistic does not imply this. The correlation, R, are higher, with a calculated 0.8 or higher in all sites.

6.2.2. Groundwater level calibration

In addition to river flow calibration, the same seven groundwater level boreholes were simultaneously used to calibrate the distributed model as in the homogenous model. The results of daily simulated groundwater levels are plotted with the observed data in Figure 6.6. The results of the calibration show wide ranging performance between sites (correlation ranging from 0.22 – 0.94), but with the general skill of the model being much poorer than the homogenous model calibration of the same groundwater levels.

In theory, the distributed model should be able to simulate both river flows and groundwater levels with greater ability, as the parameter values are in theory more refined to the particular part of the catchment and more physically realistic. However, as already noted, the practice of manually calibrating this distributed model proved difficult because of the increased number of parameters included within it.

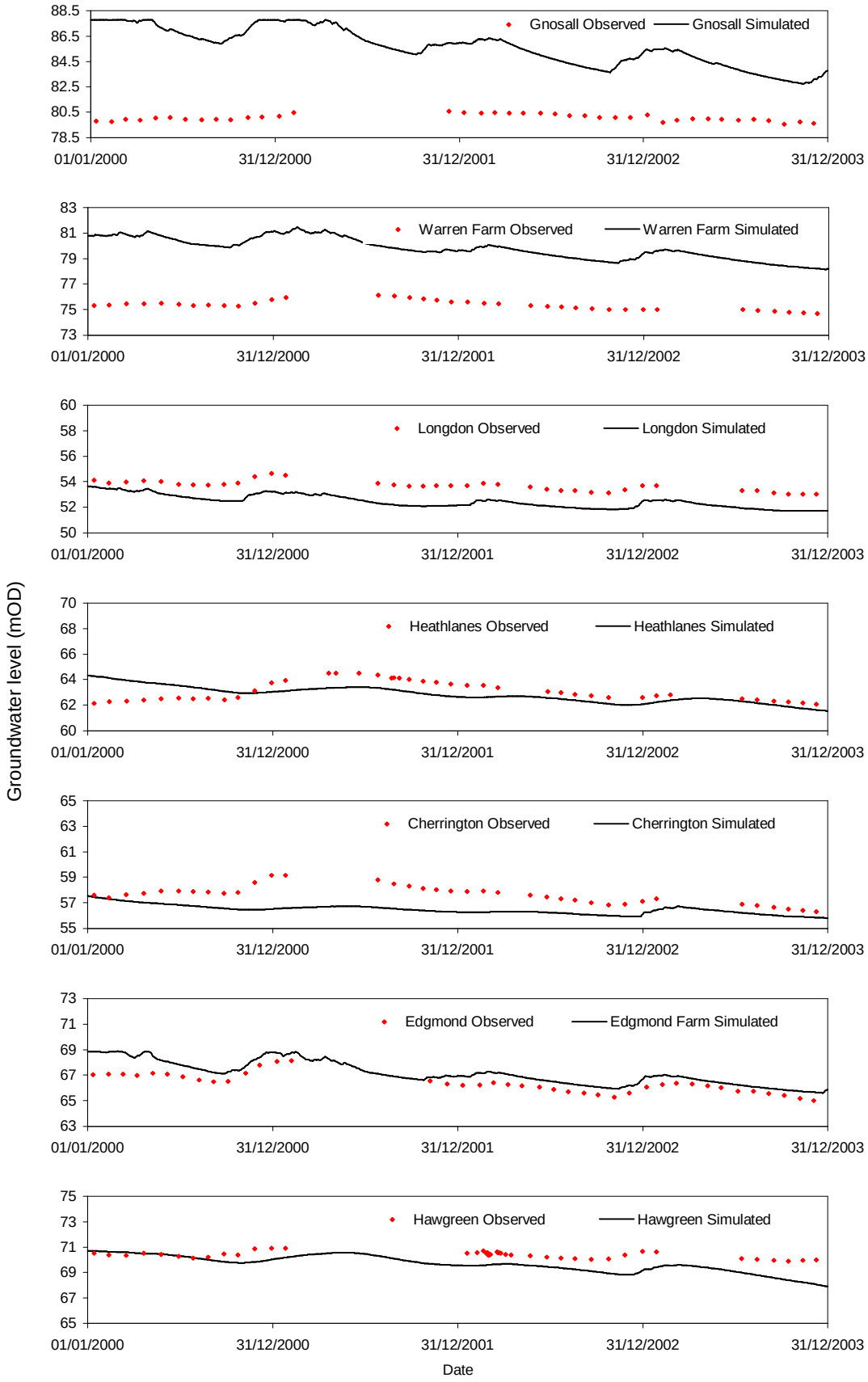


Figure 6.6. Manual calibration of groundwater levels for the distributed model

Table 6.4. Summary score for the groundwater level boreholes used in manual calibration of the distributed model

	Cherrington	Edgmond	Warren	Gnosall	Hawgreen	Heathlanes	Longdon
General level	0.3	0.36	0.02	0.01	0.3	0.32	0.28
Overall shape	0.2	0.38	0.15	0.01	0.18	0.2	0.3
Total	**0.5**	**0.74**	**0.17**	**0.02**	**0.48**	**0.52**	**0.58**

Borehole mean summary score= 0.43

The general simulated levels, shown in Figure 6.6 are not within the ~2m level that was achieved with the calibration of the homogenous model, and for this reason the general level scores within the summary score (Table 6.4) have been reduced accordingly. Once again, the level at Gnosall (close to the catchment boundary) is simulated very poorly (with a mean 5.49m general level error, Table 6.5), as well as at Warren Farm (mean 4.33m error, Table 6.5, compared to 1.17m in the homogenous manual calibration). Despite this, some of the other sites such as Longdon, Heathlanes, Hawgreen and Edgmond are simulated with levels similar to the observed data, where the model appears to show some ability. The mean summary score for these sites is slightly reduced but still comparable; from 0.695 in the homogenous manual calibration, to 0.66 for this distributed model calibration.

The coefficient of determination is calculated to quantify the relationships for the observed data and manual calibration of the distributed model in Figure 6.7 where it is seen that the R^2 is highest for the groundwater levels at Edgmond (0.874). Additionally, the closeness of the regression line to the line x=y at Edgmond indicates the ability of the model to simulate the groundwater level close to the observed level. The simulation at Longdon was very good in the homogenous model where the summary score was rated as 0.86 (Table 5.8), here the coefficient of determination between observed and simulated levels is still good, 0.740, but the plot in Figure 6.7 summarises the under-simulation in the general level, and the reduced summary score of 0.58 therefore reflects this.

Although the simulated groundwater levels in Figure 6.6 do not to fluctuate at a high temporal resolution (when compared to the daily plots of output river flow), there still appears a simulated general seasonal fluctuation that is characteristic of observed groundwater levels. In this model calibration, the peak recharge event in the period during the winter 2000-2001 is not simulated to the same extent as seen in the observed data – with only the Edgmond, Longdon and Hawgreen simulated groundwater levels appearing to display the correct seasonal variation.

Aside of the difficulty in model calibration, the reduced ability of the model in simulating groundwater levels may be due to the true spatial variability and extents of modelled parameters within the catchment being much greater and more complex (especially in the Tern catchment) than at the 1km gridded scale in the model. Despite being balanced as best possible, the parameter values that have been used are not based on specific field measurements, but estimates taken predominantly from literature. For example, Christiaens and Feyen, (2000) showed the improvement of model performance with inclusion of measured soil hydraulic characteristics when compared to the model performance with parameter values from the literature. Additionally, the discussion of issues in distributed modelling (Section 2.3) suggested that problems such as the representation of physical processes at the grid cell often limit the application and results of distributed models (such as how the Tern model has now been spatially distributed).

Table 6.5. Summary manual calibration statistics for distributed model groundwater levels (for dates where observations exist for a comparison)

	Observed mean	**Simulated mean**	**Difference (m)**	**Correlation R**	**RMSE (m)**
Longdon	53.67	52.41	-1.26	0.86	1.30
Heathlanes	63.11	62.88	-0.23	0.22	0.91
Hawgreen	70.39	69.57	-0.82	0.56	0.96
Warren	75.35	79.69	4.33	0.65	4.38
Edgmond	66.35	67.10	0.75	0.94	0.84
Cherrington	57.62	56.43	-1.20	0.49	1.35
Gnosall	80.07	85.56	5.49	0.25	5.69

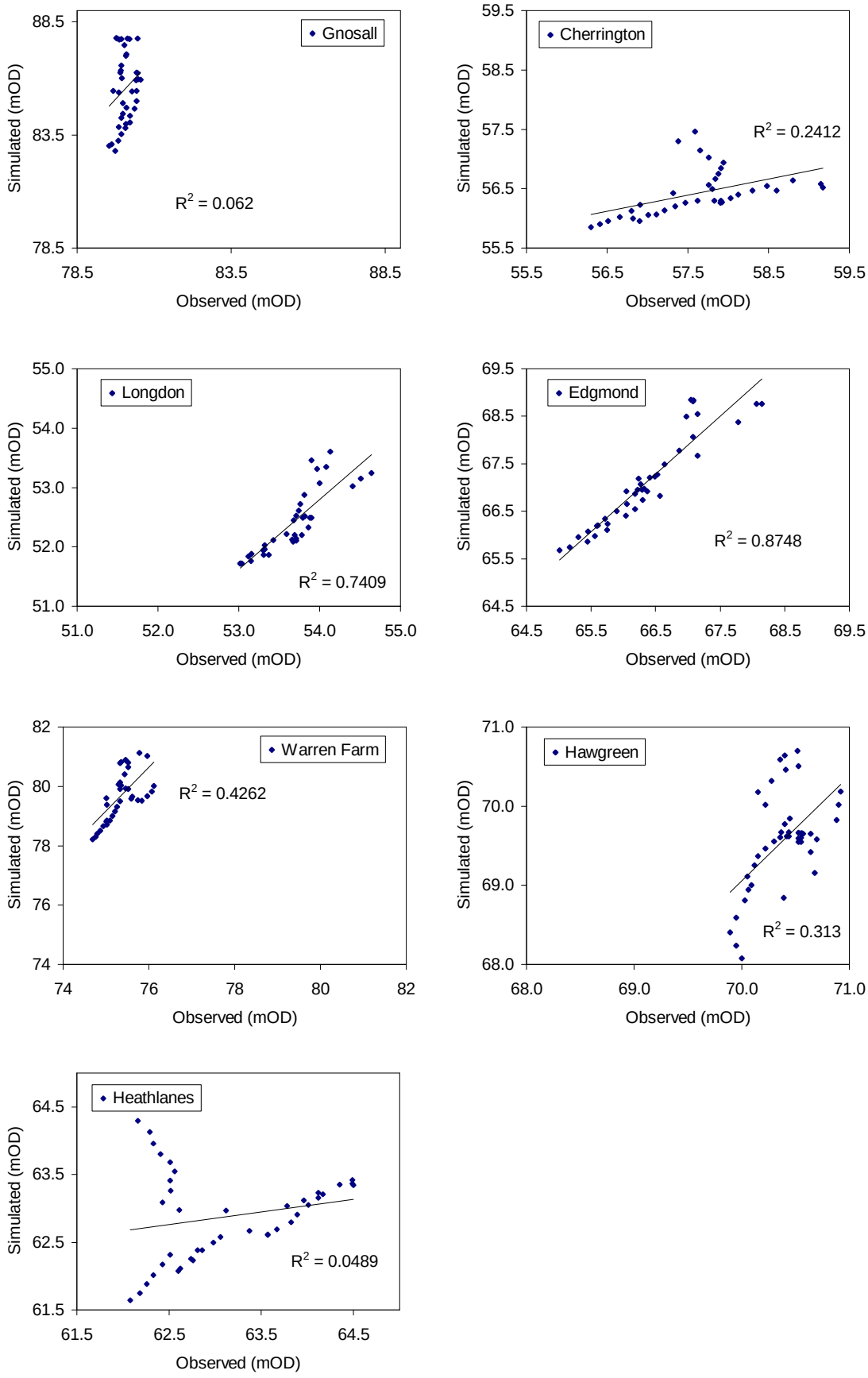


Figure 6.7. Relationships between manually calibrated groundwater levels and observed data for the distributed model

6.2.3. Validation of flows from the initial manual calibration

In general, the initial manual calibration of the distributed model simulates river flows with more ability than groundwater levels, especially in comparison to the homogenous model manual calibration. The manual calibration of the distributed model is again tested in the same way as the homogenous model in Section 5.3.3. The model outputs are presented and discussed for un-calibrated sites.

Figure 6.8 displays the output daily river flows at the four un-calibrated gauging stations where similar simulation ability is seen as the homogenous model, with Eaton-on-Tern, Tern at Ternhill and the Strine showing reasonable simulations when compared to the observed data. A much poorer simulation of flow is again seen at Potford Brook.

It is apparent that the model over-simulates flow at Tern-at-Ternhill and at Eaton-on-Tern, with higher peaks frequently simulated (especially during summer months) than observed. This over-simulation is also confirmed in the annual regime plots (Figure 6.9) where for every month the simulations are higher than observed. Despite this, the shape of the annual regimes are simulated well at these two sites, also confirmed with good correlation statistics of 0.89 (Eaton) and 0.85 (Ternhill). The NSE statistics are perhaps lower at these two sites than the R due to the noted over-simulation (the R does not consider the differences in volume, only the relationship). This is also supported with the mean daily river flows being over-simulated – 2.37 m^3s^{-1} at Eaton compared to the observed 1.95 m^3s^{-1}, and 1.42 m^3s^{-1} at Ternhill compared to the observed 1.05 m^3s^{-1}.

In contrast, the simulated flows in the Strine and Potford Brook are comparably under-simulated when compared to the observed data (Figure 6.9). This is supported by the comparisons of mean daily flows (Table 6.6) where mean daily flow in Potford Brook is 0.03 m^3s^{-1} compared to the observed 0.16 m^3s^{-1} which is significantly under represented. In Chapter 5 it was suggested this under-representation of flow that was also apparent in the homogenous model may be due to an un-representative catchment uniform value of horizontal hydraulic conductivity in a very sandstone dominated part of the catchment. However, with more specific value of HK used in this distributed model it is further confirmed that the cause is likely due to the scaling issues also suggested. The NSE for Potford Brook is -0.18 which is of comparable performance to the NSE of -0.32 in the manual calibration of the homogenous model.

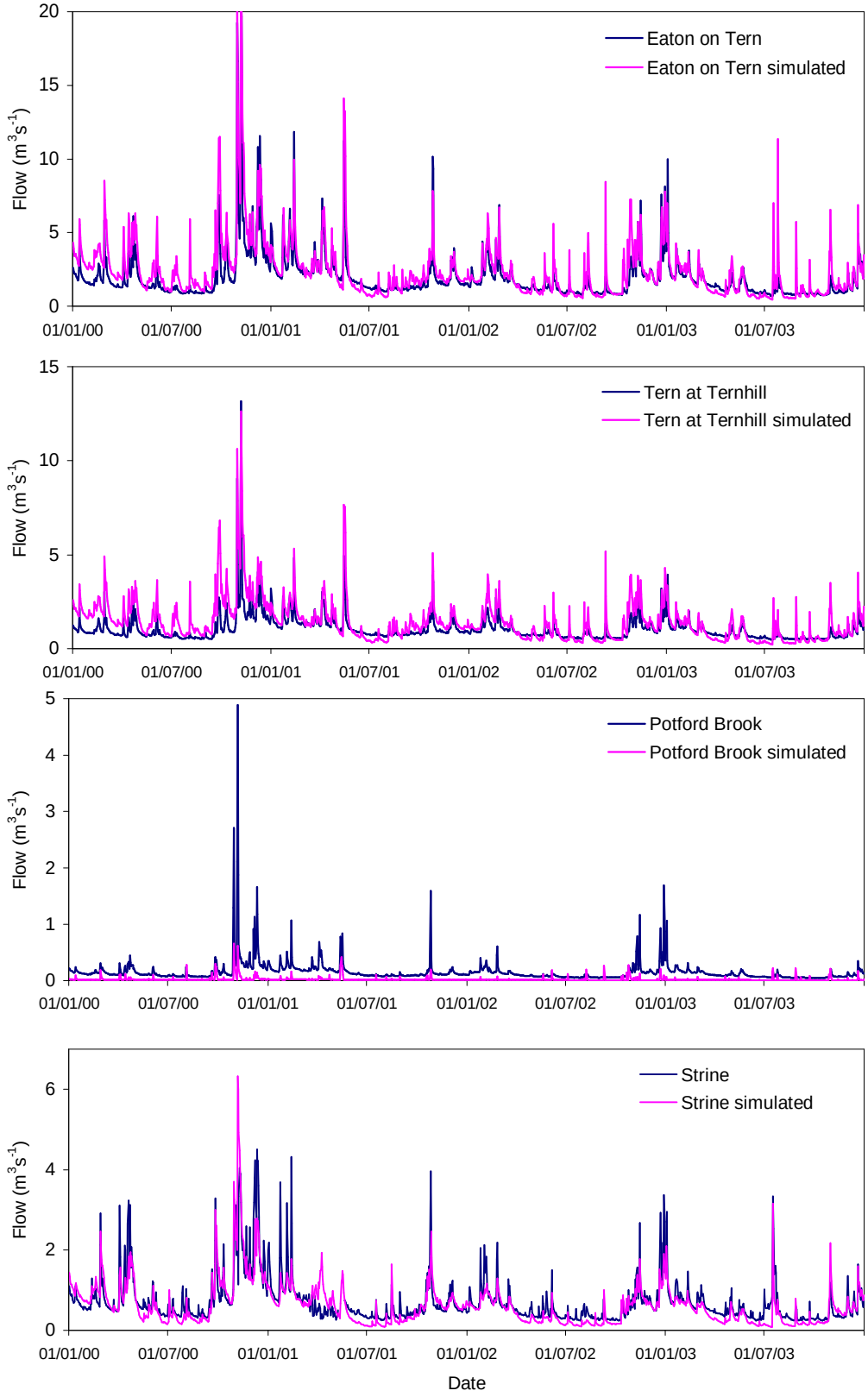


Figure 6.8. Observed and simulated daily river flows at four gauging stations used in the model testing and validation of the distributed model

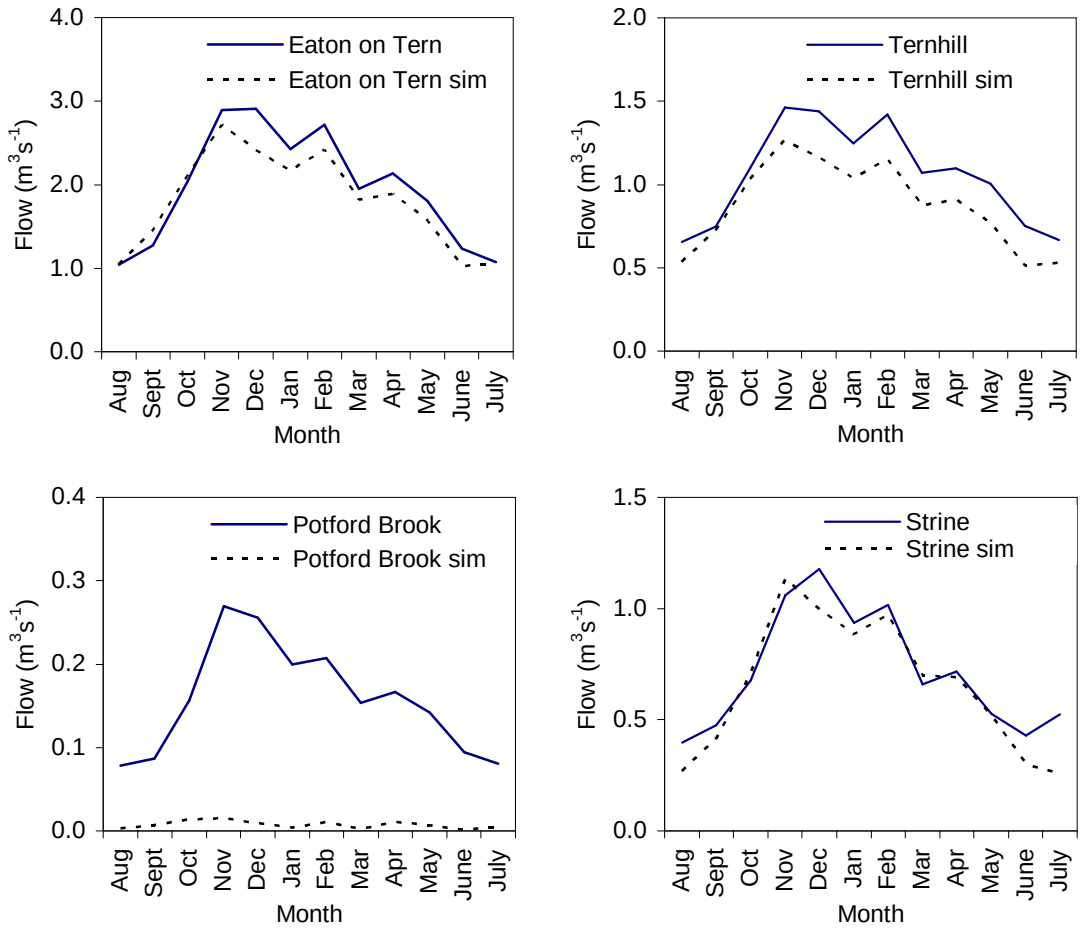


Figure 6.9. Observed and simulated annual river flow regimes used in model testing / validation for the distributed model

Table 6.6. Summary statistics for the four river flow gauging stations used in the testing and validation of the distributed model. (Observed values in bold, simulated values in normal text)

Gauge	**MDF** m^3s^{-1}	**Max flow** m^3s^{-1}	**Total flow volume** m^3	**R**	**NSE**	**RMSE** m^3s^{-1}	**RMSEP**
Eaton on Tern	**1.95**	**20.79**	2.46E+08	0.89	0.58	1.05	0.54
	2.37	24.23	3.00E+08				
Tern at Ternhill	**1.05**	**13.17**	1.33E+08	0.85	-0.16	0.74	0.70
	1.42	12.64	1.79E+08				
Potford Brook	**0.16**	**4.89**	1.99E+07	0.64	-0.18	0.23	1.44
	0.03	0.66	3.20E+06				
Strine at Crudgington	**0.72**	**4.51**	8.95E+07	0.79	0.58	0.38	0.53
	0.62	6.33	7.89E+07				

6.2.4. Groundwater level validation

Following the same method as in Chapter 5 for the testing and validation of the homogenous model manual calibration, the output simulations from the same un-calibrated seven groundwater boreholes are assessed. Figure 6.10 shows the output daily groundwater levels with the observed level data also shown for a comparison to assess model ability.

What is most apparent in the plots is that the simulated shapes of the groundwater levels in most cases show little seasonal fluctuation. For example, in the manual calibration of the homogenous model, the simulation at Woodlands Farm was very good (both the general level and the seasonal fluctuation), whereas in Figure 6.10 it is shown that the simulation is relatively flat and unvarying. To reflect this, the RMSE (Table 6.7) has increased to 1.36 m at this site; where as the manual calibration for the homogenous model was 0.52.

It is also interesting to note that the simulation at Twinney Lane (Figure 6.10) also shows little seasonal variation. When compared to Figure 5.8 for the homogenous model, there is notable reduction in the simulation ability (although the general level was not accurate in the homogenous model, the seasonal variability was comparably good). The plots therefore indicate that the homogenous model shows a better simulation at this site. However, the RMSE statistic was calculated as 1.89 m for the homogenous model, yet the RMSE for the distributed manual calibration at this site is 0.89 m which would suggest this calibration is better. The importance of assessing both the statistics and the plots is therefore highlighted in this case as the best statistics do not always result in the best fitting models.

With regard to the ability of the model to simulate the general groundwater level at these uncalibrated boreholes it is fair to suggest that the general levels are within the right magnitude, with Table 6.7 showing the largest mean level difference to be at Coley borehole, with a 2.12m level difference. As suggested in Section 5.3.4, where similar differences in groundwater levels were seen, such magnitudes of error are typical within integrated hydrological modelling studies (Refsgaard, 1997; Madsen, 2003).

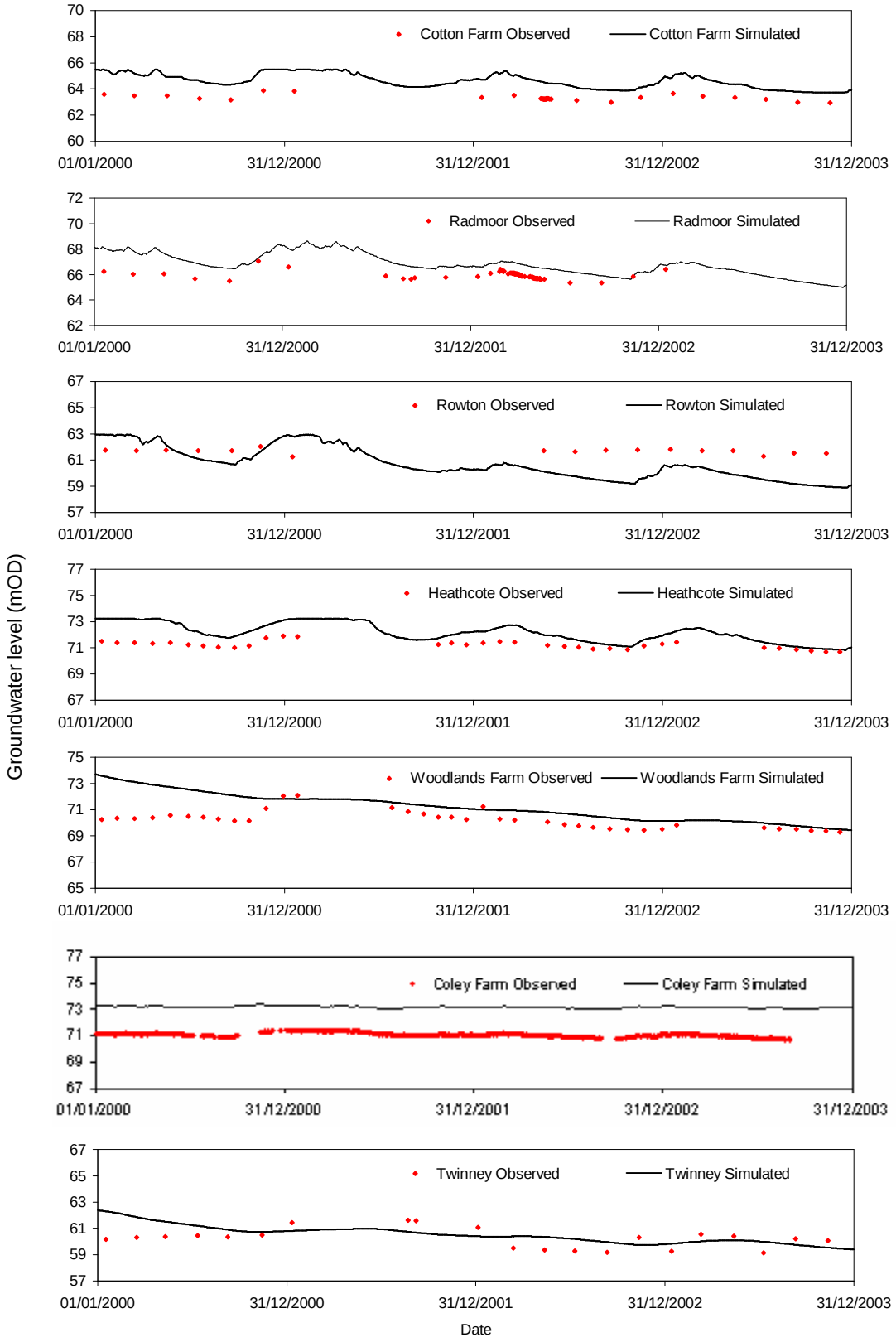


Figure 6.10. Observed and simulated groundwater levels at the seven boreholes used in the model testing and validation phase for the distributed model

Table 6.7. Summary groundwater level performance statistics for the seven boreholes used in the testing and validation of the distributed model

	Observed mean	Simulated mean	Difference (m)	RMSE (m)
Woodlands	70.21	71.22	1.01	1.36
Radmoor	65.94	66.81	0.88	0.92
Heathcote	71.25	72.07	0.86	1.00
Rowton	61.61	60.68	-1.00	1.64
Cotton Farm	63.33	64.56	1.24	1.27
Twinney Lane	60.16	60.52	0.28	0.89
Coley	71.04	73.17	2.12	2.13

To summarise, this section has described the manual calibration and validation of the distributed model for river flows and groundwater levels at various sites within the catchment. Although the calibration is fair and in some sites good, the performance of the model is not as good as for the homogenous model that was described in Chapter 5. This is likely due to the difficulty in manual calibration of a model with many more parameters. It is also apparent that where the homogenous model had difficulty in simulation at particular sites (such as Potford Brook river flow, and Gnosall groundwater level) the same problems are seen in the distributed model. This leads to a suggestion that the problem lies not with uncertainty in the type or method of calibration, but rather in the model conceptualisation (such as the grid size and the lack of sub-surface groundwater flow information close to the surface catchment boundary). The manual calibration is now used in an automatic sensitivity analysis of model parameters.

6.3. Parameter sensitivity analysis

This section includes a sensitivity analysis undertaken within the Autocal component of MIKE ZERO. Using the parameter values derived from the manual calibration of the distributed model described in Section 6.2, a test of 61 model parameters from the unsaturated and saturated zone are assessed. The aims of the sensitivity analysis are:

1) To report which parameters are most sensitive in causing variations to output flow and groundwater level RMSE when perturbed.

2) To use the results to develop a method of selecting parameters that are then taken forward for use in automatic calibrations of the distributed model.

6.3.1. Methodology

In contrast to the manual sensitivity analysis methodology adopted for the homogenous model in Chapter 5, the distributed model includes a much larger number of model parameters and it was considered unfeasible to undertake a sensitivity analysis using the same manual method. Within the MIKE ZERO Autocal component, an option of parameter sensitivity analysis allows a systematic approach to testing the sensitivity of many different parameters. In order to assess how the distributed model parameters influence river flows and groundwater levels, all 61 model parameters from the unsaturated and saturated zone have been tested. Tables 6.8 and 6.9 summarise the different parameters that have been included, and highlight the feasible parameter space ranges for each parameter with the minimum, starting and maximum allowed values shown.

The 21 unsaturated zone parameters shown relate to the upper and lower parameter limits for the distributed model setup in Table 4.10. For each of the five different soil classes, the soil water content at saturated (SAT), field capacity (FC) and field wilting point (FWP) are defined, as is the infiltration rate. The combination of these parameter values determines soil water conditions in the unsaturated zone for each class. The depth of ET parameter is also included in the sensitivity analysis as a uniform parameter and has been included as no observed data are available for the catchment for this parameter.

The saturated zone parameters that have also been spatially distributed in the model cover two layers. A layer representing the extensive drift deposits across much of the catchment (D), and a lower layer representing the solid geology (S). As shown in Table 6.9, the horizontal (HK) and vertical (VK) hydraulic conductivity, the specific yield (SY) and the storage coefficient (SC) are associated with each drift and solid geology type. In total there are 40 parameters associated with the saturated zone in the distributed model and these are all included in the sensitivity analysis.

Table 6.8. Parameters and ranges for sensitivity analysis of the distributed model in the unsaturated zone (unless noted, units are fractions)

Parameter ID	Parameter name	Initial value	Lower limit	Upper limit
SAT 1	SWC in UZ class 1, at saturated conditions	0.455	0.386	0.549
FC 1	SWC in UZ class 1, at field capacity	0.38	0.29	0.456
FWP 1	SWC in UZ class 1, at field wilting point	0.231	0.14	0.309
INF 1	Infiltration rate in UZ class 1 (m s^{-1})	5.3 e-8	4.51e-8	6.37e-8
SAT 2	SWC in UZ class 2, at saturated conditions	0.434	0.345	0.526
FC 2	SWC in UZ class 2, at field capacity	0.399	0.265	0.425
FWP 2	SWC in UZ class 2, at field wilting point	0.187	0.117	0.269
INF 2	Infiltration rate in UZ class 2 (m s^{-1})	5e-8	4.05e-8	6.13e-8
SAT 3	SWC in UZ class 3, at saturated conditions	0.442	0.378	0.546
FC 3	SWC in UZ class 3, at field capacity	0.277	0.145	0.37
FWP 3	SWC in UZ class 3, at field wilting point	0.101	0.034	0.154
INF 3	Infiltration rate in UZ class 3 (m s^{-1})	5.1 e-8	4.5e-8	6.37e-8
SAT 4	SWC in UZ class 4, at saturated conditions	0.46	0.354	0.592
FC 4	SWC in UZ class 4, at field capacity	0.325	0.098	0.531
FWP 4	SWC in UZ class 4, at field wilting point	0.162	0.026	0.322
INF 4	Infiltration rate in UZ class 4 (m s^{-1})	5.3 e-8	4.05e-8	6.83e-8
SAT 5	SWC in UZ class 5, at saturated conditions	0.434	0.372	0.535
FC 5	SWC in UZ class 5, at field capacity	0.255	0.105	0.394
FWP 5	SWC in UZ class 5, at field wilting point	0.09	0.028	0.171
INF 5	Infiltration rate in UZ class 5 (m s^{-1})	5e-8	4.28e-8	6.25e-8
ET DEPTH	Depth of evapotranspiration in UZ (uniform, m)	1.5	0.5	2.5

Where class1 = Slowly permeable soils with prolonged seasonal water-logging over slowly permeable substrates, class 2 = Slowly permeable soils with slight or seasonal water-logging over slowly permeable substrates, class 3 = Relatively permeable and free draining soils on permeable substrate with deep groundwater, class 4 = Soils with shallow groundwater (within 1m depth) and artificial drainage and class 5 = Permeable soils that are free draining and on a permeable substrate with deep groundwater (previously summarised in Table 4.11).

Table 6.9. Parameter ranges for sensitivity analysis of the distributed model in the saturated zone. *Horizontal & vertical hydraulic conductivity (m s^{-1}), specific yield & storage coefficient (fractions).

ID	Parameter name*	Initial value	Lower limit	Upper limit
D1 HK	Horizontal hydraulic conductivity in class 1 drift deposits	5e-5	1e-5	9e-5
D1 VK	Vertical hydraulic conductivity in class 1 drift deposits	9e-6	9e-9	5e-5
D1 SY	Specific yield in class 1 drift deposits	0.16	0.1	0.3
D1 SC	Storage coefficient in class 1 drift deposits	0.0001	1e-5	0.25
D2 HK	Horizontal hydraulic conductivity in class 2 drift deposits	5e-5	1e-5	9e-5
D2 VK	Vertical hydraulic conductivity in class 2 drift deposits	9e-6	9e-9	5e-5
D2 SY	Specific yield in class 2 drift deposits	0.44	0.1	0.6
D2 SC	Storage coefficient in class 2 drift deposits	0.0001	1e-5	0.25
D3 HK	Horizontal hydraulic conductivity in class 3 drift deposits	0.00028	0.0001	0.00057
D3 VK	Vertical hydraulic conductivity in class 3 drift deposits	9e-6	9e-9	5e-5
D3 SY	Specific yield in class 3 drift deposits	0.25	0.1	0.3
D3 SC	Storage coefficient in class 3 drift deposits	0.0001	1e-5	0.25
D4 HK	Horizontal hydraulic conductivity in class 4 drift deposits	5.78e-8	1e-8	1e-7
D4 VK	Vertical hydraulic conductivity in class 4 drift deposits	9e-6	9e-9	5e-5
D4 SY	Specific yield in class 4 drift deposits	0.12	0.1	0.3
D4 SC	Storage coefficient in class 4 drift deposits	0.0001	1e-5	0.25
S1 HK	Horizontal hydraulic conductivity in class 1 solid geology	9e-6	1e-8	5e-6
S1 VK	Vertical hydraulic conductivity in class 1 solid geology	2e-8	1e-11	5e-8
S1 SY	Specific yield in class 1 solid geology	0.3	0.1	0.35
S1 SC	Storage coefficient in class 1 solid geology	0.001	1e-5	0.3
S2 HK	Horizontal hydraulic conductivity in class 2 solid geology	7e-5	5e-5	9e-5
S2 VK	Vertical hydraulic conductivity in class 2 solid geology	2e-8	1e-9	5e-7
S2 SY	Specific yield in class 2 solid geology	0.25	0.1	0.35
S2 SC	Storage coefficient in class 2 solid geology	0.001	1e-5	0.3
S3 HK	Horizontal hydraulic conductivity in class 3 solid geology	2.08e-5	1e-5	9.72e-5
S3 VK	Vertical hydraulic conductivity in class 3 solid geology	1e-7	4.86e-8	1.04e-5
S3 SY	Specific yield in class 3 solid geology	0.242	0.1	0.35
S3 SC	Storage coefficient in class 3 solid geology	0.001	1e-5	0.3
S4 HK	Horizontal hydraulic conductivity in class 4 solid geology	3e-5	1.5e-5	9.72e-5
S4 VK	Vertical hydraulic conductivity in class 4 solid geology	8.5e-6	4.86e-8	2e-5
S4 SY	Specific yield in class 4 solid geology	0.232	0.1	0.35
S4 SC	Storage coefficient in class 4 solid geology	0.001	1e-5	0.3
S5 HK	Horizontal hydraulic conductivity in class 5 solid geology	8.82e-5	2e-5	1.736e-4
S5 VK	Vertical hydraulic conductivity in class 5 solid geology	2e-7	5.68e-8	2.26e-5
S5 SY	Specific yield in class 5 solid geology	0.232	0.1	0.35
S5 SC	Storage coefficient in class 5 solid geology	0.001	1e-5	0.3
S6 HK	Horizontal hydraulic conductivity in class 6 solid geology	2.08e-5	1e-5	9.72e-5
S6 VK	Vertical hydraulic conductivity in class 6 solid geology	1e-7	5.78e-8	1.04e-5
S6 SY	Specific yield in class 6 solid geology	0.265	0.1	0.35
S6 SC	Storage coefficient in class 6 solid geology	0.001	1e-5	0.3

Where: D1= Till/diamicton, D2= peat, D3= glacio-fluvial sands and gravels, D4= alluvial silty clay, S1= low permeability mudstones, S2= mixed permeability class, S3= Bromsgrove/Helsby Formation, S4= Kidderminster/Chester Pebble Beds, S5= Bridgnorth/Kinnerton Formation, S6= Wildmoor/Wilmslow Formation

Parameter sensitivity on output river flow and groundwater levels are assessed using the aggregation of two objective functions, the weighted sum of squares for the RMSE of all eight gauging stations, and the weighted sum of squares for the RMSE of all 14 groundwater boreholes (a total of 22 output measures). This approach of assessing sensitivity based on all the observed data available results in parameter sensitivity being assessed in all sub-basins and from different (although not all) types of geology. In addition to the aggregate objective function, the individual output measures have also been recorded in order to separately better understand which parameters influence which sub-basins river flows and groundwater levels.

The parameter sensitivity is assessed in a central difference approach of a 5% fraction of the parameter interval. The methodology assesses the RMSE for each output measure for an initial run (1) and subsequently 61 model runs in a forward perturbation (increasing the initial parameter value) by 5% of the parameter range, and then a further 61 model runs varying each parameter in a backward perturbation (decreasing the initial parameter value) by 5% of the allowed range. In total 123 model simulations were undertaken to assess the sensitivity of the different parameters, covering a 10% range from the initial parameter value defined in Tables 6.8 and 6.9.

6.3.2. Results

The results of the sensitivity analysis are shown in Figures 6.11 for river flows and 6.12 for groundwater levels. The plots show individual results for each parameter perturbation at each gauging station or borehole and are expressed as the change in the RMSE value for each of the 5% perturbations in the forward and backward directions. The left part of each figure shows the results for the unsaturated zone parameters, and the right side shows the saturated zone parameters. Red threshold bands have also been added at +/- 0.5% change from the initial RMSE value for each site as an indication of a degree of sensitivity above which the perturbation causes a significant variation in the output flow or groundwater level.

6.3.2.1. River flows

As should be expected with the distributed model, Figure 6.11 indicates that there are individual sensitivity profiles for each gauging station, where no two gauging stations

respond in the same way to parameter variation. With the spatial distribution of the model, different parameters are present within different sub-basins and to varying extents. For example, it is shown in the unsaturated zone that variation to classes 1 and 3 (especially the SAT parameter), result in relatively large changes to the RMSE but at differing extents depending on the site. Eaton-on-Tern and Tern at Ternhill are located in areas of the catchment where the upstream soil classification (3) (Table 4.11) is a permeable free draining soil on permeable substrate with deep groundwater (the sandstone aquifers) and for this reason this parameters variation results in an impact on these nearby river flows.

The ET depth parameter also in the unsaturated zone is the only model parameter in the sensitivity analysis that is not spatially distributed. Figure 6.11 indicates that its variation results in moderate changes to the RMSE at all sites except Potford Brook. In the homogenous model this was also listed as an important model parameter and so it was expected that it should still be influential on model outputs.

It is shown that perturbation of many of the drift model parameters (D) in the saturated zone result in smaller changes to outputs than the solid (S) SZ parameters. This is somewhat expected as the drift deposits do not completely cover the whole area of the catchment and in parts are fragmented (Figure 3.6) and therefore likely to have less of an impact on output flow RMSE. Where substantial drift deposits are located, however, in the Roden and Bailey Brook to the north-west of the catchment, it is shown that the D1 HK (till/diamicton) and D3 HK (glaciofluvial sands and gravel) parameters are more influential.

Perturbations of the horizontal hydraulic conductivity solid geology parameters in the saturated zone, notably S1 (low permeability mudstones) and S4 (sandstone, Kidderminster/Chester Pebble Beds) are also shown to result in significantly large (greater than 0.5%) changes to output flow RMSE. These two classes are also the two largest by area covering the catchment. It is expected that the horizontal hydraulic conductivity parameter should be important in influencing flow as this was shown to be the case in the homogenous model.

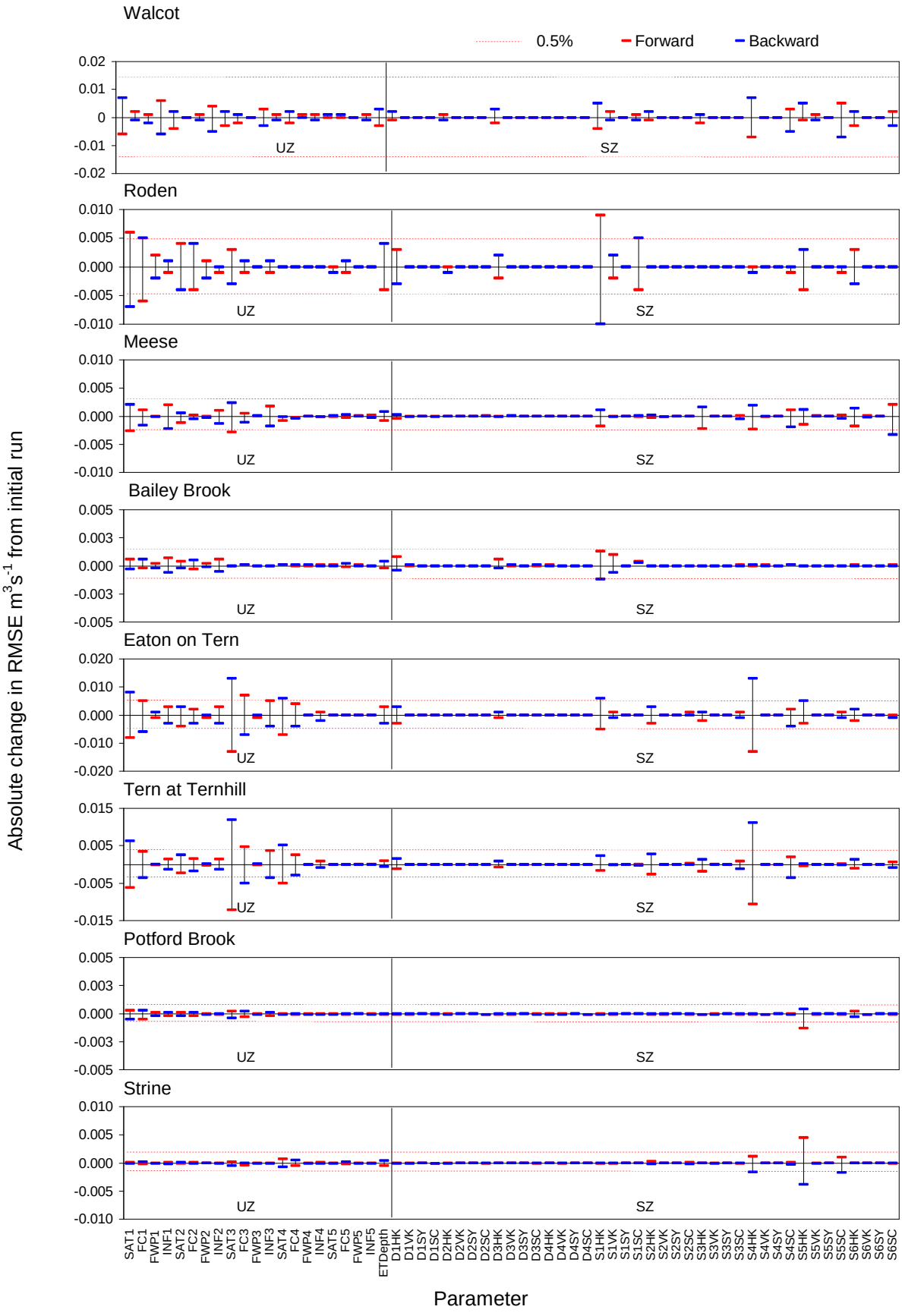


Figure 6.11. Sensitivity of model parameters at eight gauging stations after a 5% perturbation (forward and back) from the parameter interval. The dashed red line indicates the assessed threshold (0.5% from initial run) that defines whether the parameter is sensitive to model performance

As expected, Roden and Bailey Brook (and by downstream influence, Walcot) are all sensitive to variation of the S1 parameter (the low permeability mudstone class), and Eaton and Tern-at-Ternhill sensitive to the sandstone dominant class, S4.

In addition to horizontal HK, it is also important to note that the perturbations of vertical HK, specific yield and storage coefficient (that had negligible effects in the homogenous model due to there only being one saturated zone layer), now result in variations, however, these are smaller than for horizontal HK in output RMSE.

6.3.2.2. Groundwater levels

The plots shown in Figure 6.12 for groundwater level boreholes are constructed using the same method as in Figure 6.11. Compared to the simulated river flow results, it is clear that fewer parameters have an influence on output groundwater levels for the range assessed. In the unsaturated zone, the SAT and FC parameters (especially for classes 1 and 3) are influential on the output results of groundwater level RMSE. As shown in the sensitivity analysis of the homogenous model (Section 5.4), the SAT and FC parameters (whose perturbations were also the most sensitive in the homogenous model) govern the volume of water that is held within the sub-surface unsaturated zone, and therefore the result that the perturbation of these parameters influences groundwater levels is not unexpected.

As with the river flow gauging stations, perturbation of the drift classes results in little (if any) change in the output RMSE of the groundwater levels at the 14 boreholes, as also shown in Figure 6.12. Again, this is likely due to the fragmented nature of drift across the catchment and also that most of the drift is located in the north-west and not where these 14 boreholes are located. Despite this, the perturbation of the solid SZ classes, notably the horizontal hydraulic conductivity parameters result in much larger changes to the RMSE. Parameters S4HK (Kidderminster/Chester Pebble Beds), S5HK (Bridgnorth/Kinnerton Formation) and S6HK (Wildmoor/Wilmslow Formation), show the largest sensitivity on model output. Given that the boreholes are located within these sandstone classes it is not unexpected that these results are shown as the horizontal

hydraulic conductivity parameter was also shown in the homogenous model sensitivity analysis (Section 5.4) to be one of the most sensitive parameters on model outputs.

Notably, there are no data for boreholes located in S1 geology (low permeability mudstone class). However, the flow RMSE results in Figure 6.12 indicated that S1HK was an influential parameter on output flows. Therefore it is likely still an important parameter and that if this sensitivity analysis was undertaken again, it is suggested that it would be beneficial to test groundwater level outputs in all the different solid geology classes rather than solely at the borehole locations for which there was observed data. It was not possible to do this using the method adopted for this sensitivity analysis, as the RMSE output measure used to assess output changes uses the observed data within the calculation.

Figure 6.12 also highlights that, in general, the change in RMSE values are balanced for the forward and backward perturbations of the parameters. However, there is a reverse response of the model to changes in S5HK (Bridgnorth/Kinnerton Formation). Some sites demonstrate an increase to the RMSE if the perturbation is in a forward direction (increasing the HK) where as at other sites the opposite response is shown. Six of the boreholes show a decrease to the RMSE if the S5HK parameter is perturbed in a forward direction (i.e by increasing the parameter from its initial value of 6.82 e -5 m s^{-1} to 7.6 e -5 m s^{-1} or 5.89 m d^{-1} to 6.57 m d^{-1}). These reverse responses may be due to the type of geology in which the boreholes are situated. Four of these boreholes; Woodlands, Twinney Lane, Edgmond and Warren Farm are located within the same geology of the Kidderminster and Chester Pebble beds, with the other two boreholes (Radmoor and Cotton Farm) located in a geology with overlaying drift deposits.

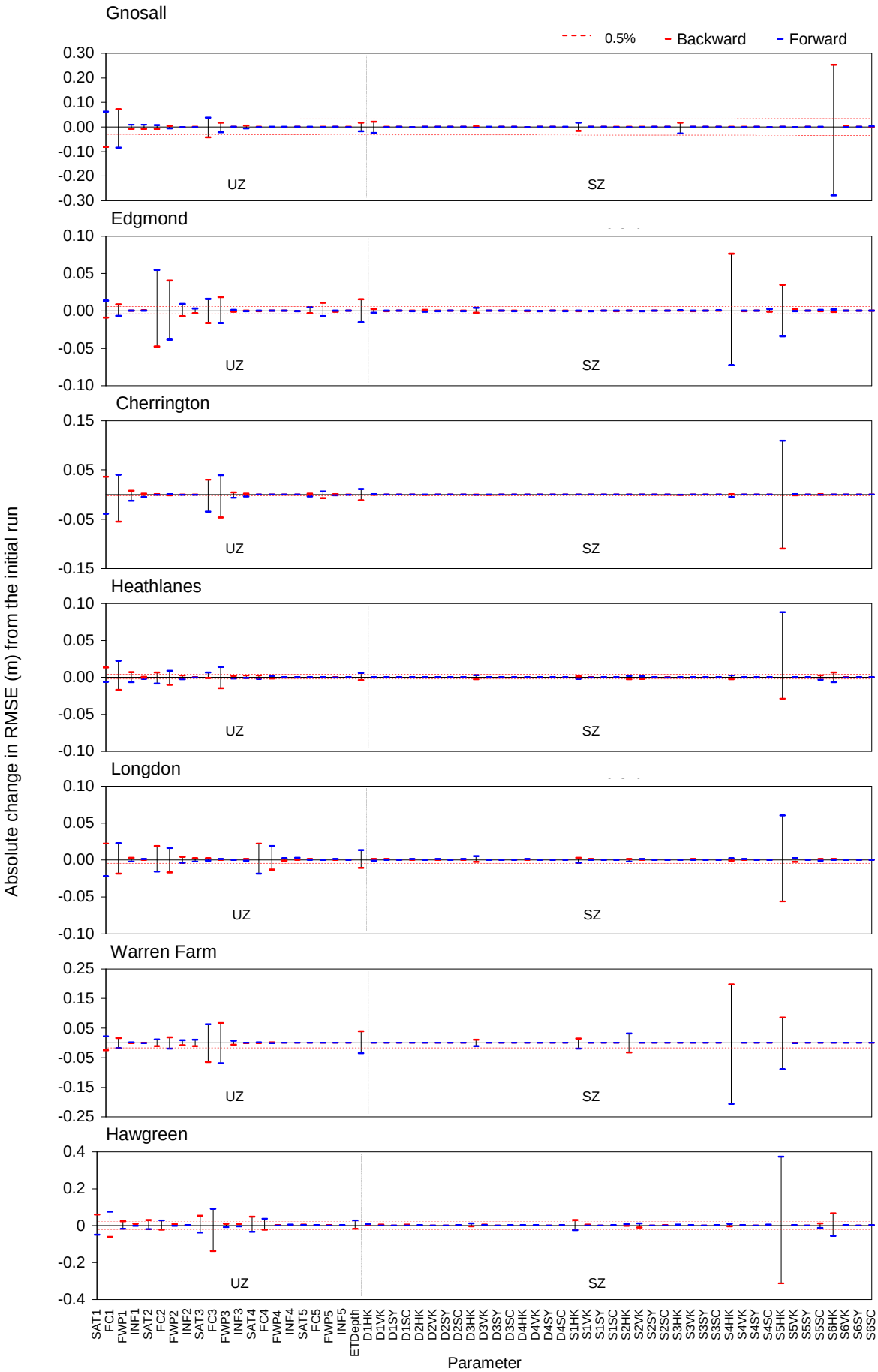


Figure 6.12. Sensitivity of parameters at 14 boreholes after a 5% perturbation (forward & back). Dashed red line indicates the threshold (0.5% from initial run) that defines whether the parameter is sensitive to model performance *(cont..)*

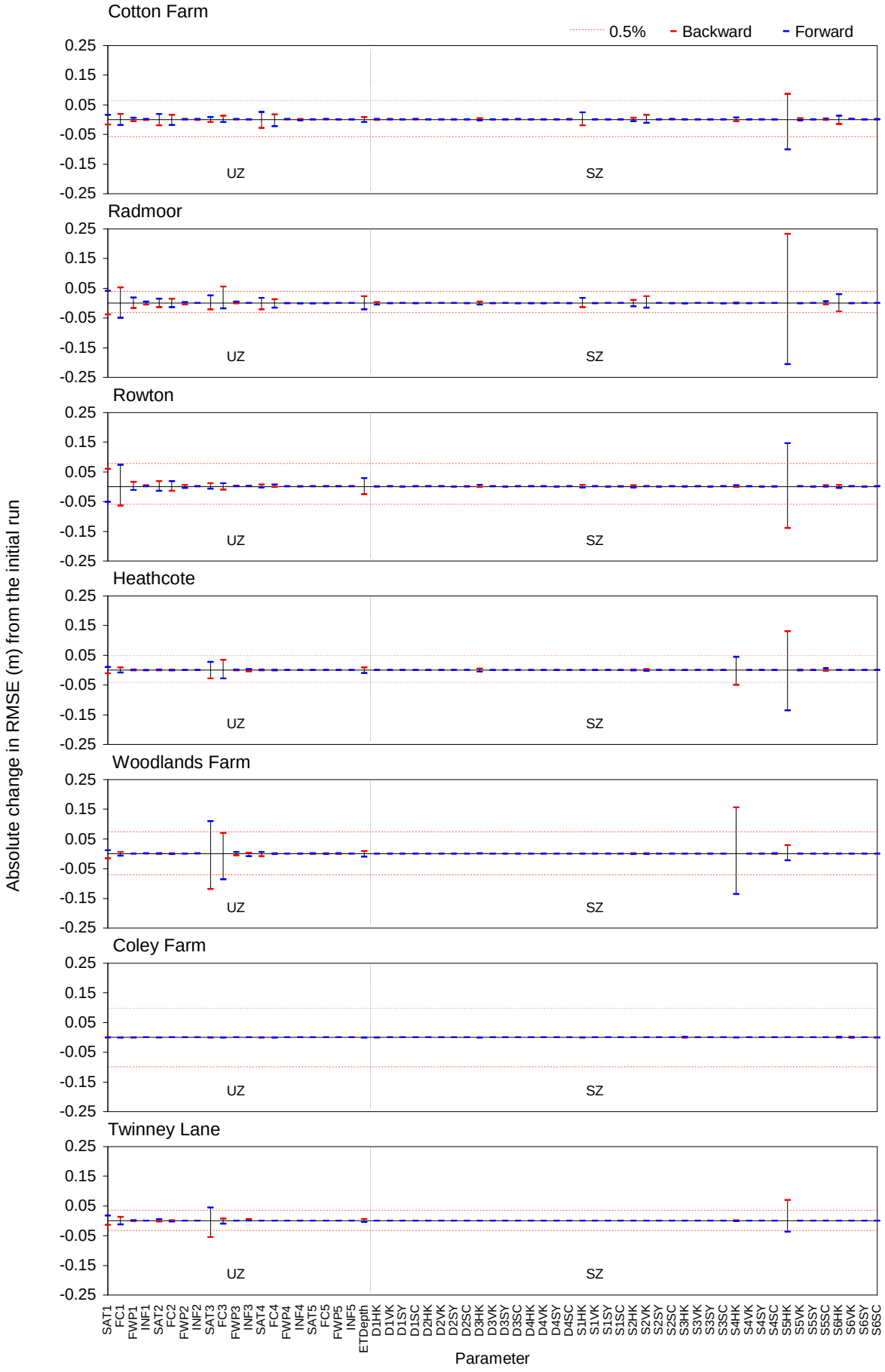


Figure 6.12. (*Cont.*.)Sensitivity of parameters at 14 boreholes after a 5% perturbation (forward & back). Dashed red line indicates the threshold (0.5% from initial run) that defines whether the parameter is sensitive to model performance

6.3.2.3. Selection of parameters

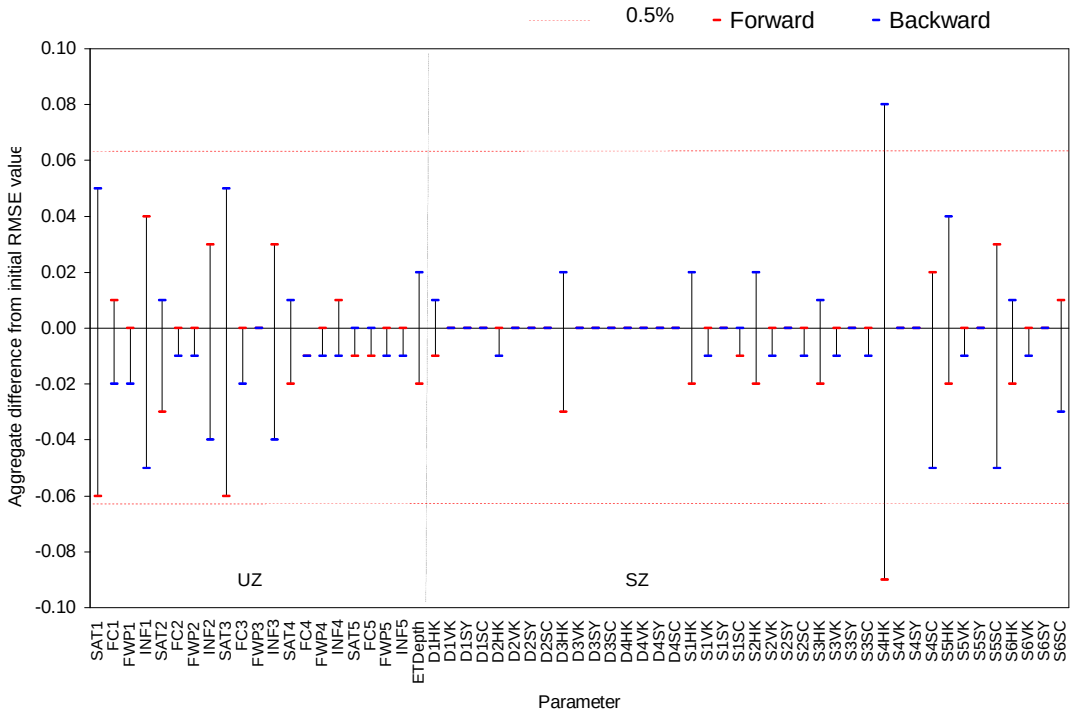
The purpose of the sensitivity analysis, aside from assessing which individual parameter perturbations result in sensitivity to models outputs, is to summarise which of these parameters are most sensitive to output flow and groundwater levels and then use these parameters in subsequent automatic calibrations. The sensitivity analysis component within MIKE ZERO reports not only the individual parameter RMSE but also the aggregate RMSE separately for both flow and groundwater levels. The aggregate measure considers all of the output measures (8 gauging stations and 14 boreholes, a total of 22) in its calculation, and is not biased towards any individual parameter. In order to maintain a similar method as the homogenous model sensitivity analysis, the aggregate has been used from which to classify if a model parameter is sensitive or not.

By selecting a threshold of a 0.5% change to the RMSE (whether this is an increase or decrease, as shown by the different forward and backward perturbations) it is possible to assess which parameters are most likely to be effective as calibration parameters whilst not over-parameterising the calibration with the inclusion of parameters whose perturbations do not result in large change to the output simulations.

The aggregate changes to flow and groundwater level RMSE are shown in Figure 6.13 a and b where the 0.5% threshold lines are also indicated. For flows this results in any change to the RMSE +/- 0.00632 m^3s^{-1} and for groundwater levels +/- 0.355 m. As in the homogenous model sensitivity analysis, the flow and groundwater classes had different thresholds due to the difference in the aggregate value from which the percentage threshold is calculated.

The plots indicate that the 0.5% change in the aggregate RMSE values results in a realistic and balanced number of model parameters to include within the automatic calibrations. Figure 6.13a highlights that for flow the S5HK parameter (Bridgnorth/Kinnerton Formation) is highlighted as the most sensitive, with many of the unsaturated zone parameters also very close to the threshold limit. The Sat 1 and 3, and Fc 1 and 3 classes are also shown in the groundwater level plot in Figure 6.13b to be above the 0.5% threshold.

(a)



(b)

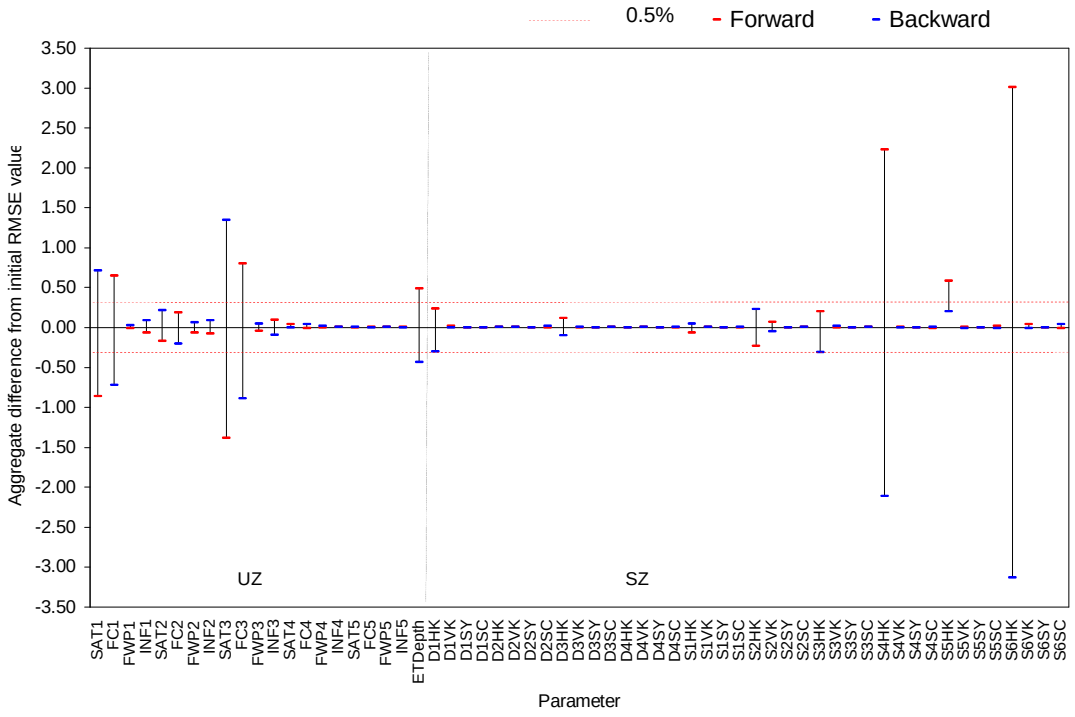


Figure 6.13. Aggregate sensitivity analysis for (a) river flows and (b) groundwater levels

Figure 6.13b also indicates that the ET depth parameter is located outside of the 0.5% threshold. In addition to the S4HK parameter, the S5HK and S6HK parameters are also highlighted above and below the threshold for groundwater levels. As already noted, the S5HK parameter showed mixed response in the forward and back direction of perturbation in Figure 6.12. The aggregate measure used in Figure 6.13 therefore

indicates that when groundwater level RMSE sensitivities are aggregated, the backward response results in a positive response when compared to the other parameters. Despite this, the parameter is still found to be outside the threshold and thus included as a sensitive parameter. This response (which may be due to locations in different geologies) highlights that the model response to parameter perturbation is not always linear.

Table 6.10 summarises the results of the sensitivity analysis, showing the eight parameters that have been selected for subsequent automatic calibration. This number of parameters is comparable to the seven parameters that were selected as a result of the homogenous model sensitivity analysis. Table 6.10 additionally indicates how many of the total 22 output measures (gauging stations and boreholes) are classified as sensitive (above the aggregate 0.5% threshold) for each parameter. The Sat 1 parameter is shown to be significantly sensitive for 16 of the 22 measures in both the forward and backward perturbation whereas only five of the measures show the S6HK parameter as significantly sensitive. Despite this, they have been included as at the individual locations where sensitivity is seen, the changes to RMSE are large.

Table 6.10. Selected parameters resulting from the sensitivity analysis (note the river flow values appear smaller as the initial aggregate flow RMSE is smaller than for groundwater levels). (If values are given in bold they are classified as outside of the 0.5% threshold)

Flow initial RMSE value = 12.64 m^3s^{-1} , 0.5% significant value 0.0632 m^3s^{-1}
GW initial RMSE value = 70.91 m , 0.5% significant value 0.355 m

Parameter ID	**Δ in aggregate Flow RMSE from initial value**		**Δ in aggregate GW level RMSE from initial value**		**Total no. measures (of 22) where > 0.5% sensitivity is recorded**	
	Backward	**Forward**	**Backward**	**Forward**	**Backward**	**Forward**
SAT 1	-0.06	0.05	**-0.86**	**0.71**	16	16
FC 1	0.01	-0.02	**0.65**	**-0.72**	12	13
SAT 3	-0.06	0.05	**-1.38**	**1.35**	13	13
FC 3	0.00	-0.02	**0.80**	**-0.89**	13	13
ET depth	-0.02	0.02	**0.49**	**-0.43**	11	11
S4 HK	**-0.09**	**0.08**	**2.23**	**-2.11**	6	8
S5 HK	-0.02	0.04	**0.59**	0.20	14	13
S6 HK	-0.02	0.01	**3.01**	**-3.13**	5	5

SAT1 = Soil water content at saturated conditions and FC1 = Soil water content at field capacity for slowly permeable soils with prolonged seasonal water-logging over slowly permeable substrates.
SAT 3 = Soil water content at saturated conditions and FC3 = Soil water content at field capacity for relatively permeable and free draining soils on permeable substrate with deep groundwater
S4 HK = horizontal hydraulic conductivity of Kidderminster/Chester Pebble Beds, S5 HK= horizontal hydraulic conductivity of Bridgnorth/Kinnerton Formation, S6 HK= horizontal hydraulic conductivity of Wildmoor/Wilmslow Formation.

Table 6.10 also indicates that parameter sensitivity is well balanced in both the forward and backward directions with a similar number of measures classed as sensitive in each direction. Additionally, the selected list also contains a balanced mix of both unsaturated and saturated zone parameters.

There are also some limitations that need to be acknowledged as a result of the method adopted for this sensitivity analysis. The test that has been used does not consider any model sensitivity to parameter perturbation outside of the range assessed (a total of a 10% perturbation from the initial calibrated value). Therefore, any non-linearity in parameter response that may be seen at other percentage parameter perturbations are not properly tested, nor are any correlations that may exist between parameters.

Because of the large number of figures that would have been required, it was also not possible to include the same output flow and groundwater level figures that were shown in the homogenous model sensitivity analysis. As has been shown elsewhere in the thesis, the assessment of output plots has been shown to aid in a better understanding of the model response. Despite this, the method that has been used provides a successful indication of the parameter sensitivity for each output measure, by providing both individual output measure plots as well as aggregate summaries.

6.3.3. Summary of sensitivity analysis

The sensitivity analysis has reported on a total of 61 model parameters using an automatic framework that employed a central difference test at 5% of the parameter range for each parameter. Percentage changes of each parameter have been calculated with a 0.5% threshold used as a limit of which to classify parameter sensitivity. Using the aggregate RMSE scores reported by the sensitivity analysis tool in MIKE ZERO, eight model parameters have been selected that are used in automatic calibrations in the subsequent section.

6.4. Automatic Calibration

Having identified model parameters that are sensitive to change in the previous section, the selected parameters are now used within the same automatic calibration framework

that was presented in Chapter 5 for the homogenous model. The section details the methodology and results of the same two automatic calibrations used with the homogenous model, one that compares a method of automatic calibration only using observed flow data at the basin outlet, and the other using the comprehensive multi-location and multi-criteria approach. The approach used for both automatic calibrations once again have the same simulation specifications, model parameters and parameter optimisation criteria with only the objective functions differing between the set ups. The purpose of these automatic calibrations are the same as those in Chapter 5:

1) To test whether the automatic calibration methodology results in an improved simulation ability of the distributed Tern hydrological model with regard to the same river flow and groundwater level statistics used in the manual calibration in Section 6.2.

2) To compare the complexity of the automatic calibration methods, and whether by including more rigorous objective functions (in the second test) the ability of the model simulations of river flow and groundwater are improved.

3) To assess and quantify the equifinality within the parameter space. This is undertaken by recognition and selection of a number of calibrated models for each automatic calibration providing a total of 20 calibrated models with different parameter values, the same number of models as was selected in the homogenous model automatic calibrations.

Using the same method as in Chapter 5, the aim is not to specifically compare the homogenous and distributed models within this chapter, rather to compare the distributed manual calibration with the distributed automatic calibrations. Chapter 7 proceeds to address these comparisons between homogenous and distributed models, as well the other research questions that were outlined in Chapter 1.

6.4.1. Methodology

As in Chapter 5, the automatic calibration for the distributed model was undertaken using the Autocal component of MIKE ZERO. The manually calibrated distributed Tern model was used as the template from which to undertake the automatic calibration. For

means of comparison, the evaluation period has been kept the same as in all model assessment in this thesis; 01/01/2000 – 31/12/2003.

Eight calibration parameters have been included in the automatic calibration that were selected as a result of the parameter sensitivity analysis in Section 6.3. Table 6.11 summarises the specified parameter ranges, the lower and upper bounds, as well as the initial value (derived from the manually calibrated model). In each simulation of the automatic calibration, the parameter set is again determined by Monte-Carlo sampling for a random test of parameter values within the defined limits. The parameter ranges have been specified according to realistic and feasible bounds defined from secondary field data and relevant literature, and attempt to be both wide enough to capture the intra-basin variation for each parameter whilst at the same time limited as best possible so that they are still realistic in nature.

Table 6.11. Initial, lower and upper parameter values used in automatic calibration of the distributed model (unless stated, units are fractions)

Parameter	**Type**	**Initial value**	**Lower bound**	**Upper bound**	**Transformation**	**Condition**
SWC SAT1	Variable	0.455	0.386	0.549	Real	
SWC FC1	Dependent	-	-	-	Real	((SWC SAT-SWC FWP)*0.8+SWC FWP)
SWC FWP1	Variable	0.231	0.14	0.309	Real	
SWC SAT 3	Variable	0.442	0.378	0.546	Real	
SWC FC 3	Dependent	-	-	-	Real	((SWC SAT-SWC FWP)*0.8+SWC FWP)
SWC FWP 3	Variable	0.101	0.034	0.154	Real	
Depth to ET surface (m)	Variable	1.5	0.5	2.5	Real	
S4 HK ($m\ s^{-1}$)	Variable	3.0 e-5	1.5 e-5	9.72 e-5	Logarithmic	
S5 HK ($m\ s^{-1}$)	Variable	8.82 e-5	2 e-5	1.736 e-4	Logarithmic	
S6 HK ($m\ s^{-1}$)	Variable	2.08e-5	1e-5	9.72 e-5	Logarithmic	

SAT1 = Soil water content at saturated conditions and FC1 = Soil water content at field capacity for slowly permeable soils with prolonged seasonal water-logging over slowly permeable substrates.
SAT 3 = Soil water content at saturated conditions and FC3 = Soil water content at field capacity for relatively permeable and free draining soils on permeable substrate with deep groundwater
S4 HK = horizontal hydraulic conductivity of Kidderminster/Chester Pebble Beds, S5 HK= horizontal hydraulic conductivity of Bridgnorth/Kinnerton Formation, S6 HK= horizontal hydraulic conductivity of Wildmoor/Wilmslow Formation.

Table 6.11 shows that the automatic calibration of the distributed model includes two dependent parameters, the soil water content at field capacity for classes SWC Fc 1 and SWC Fc 3, for the same reasons as described in Chapter 5. The parameters are again fixed to always be a fraction of 0.8 added to SWC-FWP of the interval between SWC-SAT and SWC-FWP, so that the unsaturated zone structure of the homogenous and distributed models are conceptualised in the same way, and only the spatial extents of different classes vary.

As previously introduced, automatic calibration is undertaken on two sets of simulations, one that uses only one output measure (the RMSE of flow at the basin outlet, Walcot) and another that adopts the more rigorous multi-criteria and multi-location approach with two objective functions that are weighted equally. Again the aim is to assess overall performance not to assign any particular stream or groundwater level as more important. The same objective functions are used, where one objective function sums the RMSE of river flow at four gauging stations, and the other sums the RMSE of groundwater levels at seven boreholes, as was conceptualised in Figure 5.19.

As in Chapter 5, both sets of automatic calibrations use almost exactly the same algorithmic parameters within the Shuffled Complex Evolution method of parameter optimisation. Table 6.12 summarises the particular algorithm values and the stopping criteria.

Table 6.12. Summary of Algorithmic parameters and stopping criteria used in the automatic calibration of the distributed model
(where n is the number of variable parameters, in this case, 7)

Algorithmic parameter / *stopping criteria	**Recommended value**	**Value used**
Number of complexes	-	3
Number of points in a complex	2n+1	15
Number of points in a sub-complex	n+1	8
Number of evaluation steps by each complex before shuffling	2n+1	15
*Maximum number of model evaluations	-	1000 & 4000
*Number of loops of convergence	-	3
*Minimum relative change in objective function value	-	0.01

6.4.2. Automatic calibration and testing at the basin outlet (Walcot)

Figure 6.14 shows the results of the optimisation from the automatic calibration of the distributed model using the output measure of RMSE of flow at the basin outlet, Walcot. The convergence criteria were met after 342(+1 initial) simulations of the model using different parameter sets. The optimal RMSE value for flow at Walcot gauging station that was achieved was 2.989 m^3s^{-1} (RMSEP = 0.405) as highlighted in Figure 6.14, compared to 3.07 (RMSEP = 0.42) for the manual calibration of the distributed model.

6.4.2.1. Selection of calibrated models and parameter variation

As was shown in Figure 5.20 for the homogenous model automatically calibrated using the RMSE of flows at Walcot gauging station, the plot in Figure 6.14 again identifies consistent improvement in the optimisation of the RMSE throughout the calibration process, with RMSE values contained within 2% of the optimum value (which ranges from 2.989 to 3.05 m^3s^{-1} as highlighted by the horizontal threshold line) from simulation 200 onwards.

Using the same method as for the homogenous model, the threshold of 2% from the optimal RMSE (also shown in Figure 6.14) again defines the limit for which models are considered calibrated, thus recognising the equifinality of different parameter sets producing statistically similar results. As shown by the highlighted red simulations, there are 191 models that are located beneath this threshold.

As noted in the previous chapter, Autocal does not report the Nash-Sutcliffe NSE or correlation R statistics that are calculated and used to assess model performance. To compare the performance of the automatically calibrated models to the manually calibrated models using the same statistics, a selection of a suite of ten models has also been undertaken for the distributed model, and the ten models then re-simulated within MIKE SHE in order to calculate the performance statistics.

The selection of models was again undertaken systematically to ensure a fair sample across the range of 191 calibrated models, these models are also highlighted in Figure 6.14. Selecting the optimal model (with lowest RMSE), the model at the 2% limit (RMSE = 3.05 m^3s^{-1}) and a further 8 (with every 24^{th} model selected of the remaining 189 models), ensured a representative sample.

Table 6.13 provides further detail of the specific parameter values associated with each of the ten sampled models. The table also summarises the minimum and maximum ranges of each parameter (the calibrated parameter space) as well as a comparison to the parameter space for all 191 calibrated models. As was previously shown in Chapter 5, the parameter ranges are larger for the 191 models (all models within the 2% threshold) than for the sampled set of ten models, however, the full range in RMSE at the basin outlet are included within the sample.

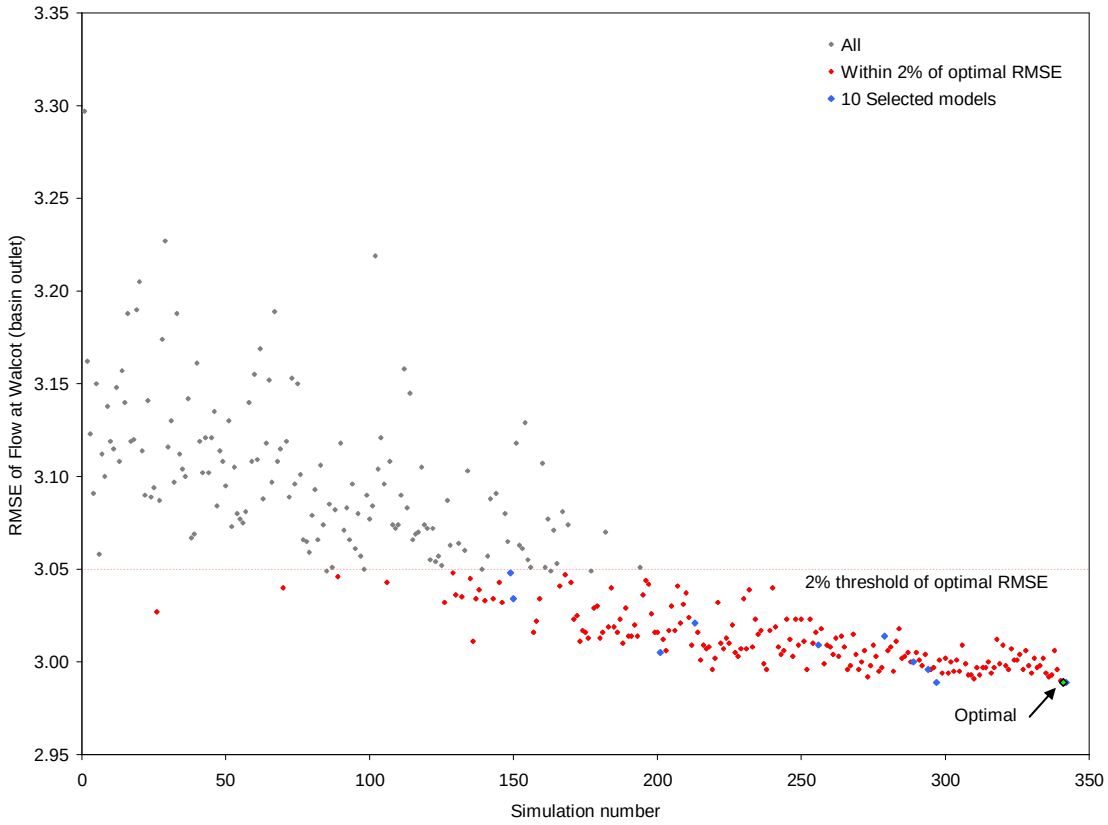


Figure 6.14. Optimisation of RMSE m^3s^{-1} at the basin outlet for the automatically calibrated distributed Tern model

The optimisation of the individual parameters are highlighted in Figure 6.15. These plots show the values of each parameter adopted with each model run. The plots display all simulations, the calibrated simulations (red) and the sampled set of ten (blue) as well as the mathematically optimum model (green). The plots indicate that during the automatic calibration process, the UZ1 class, slowly permeable soils with prolonged seasonal water-logging over slowly permeable substrates (SWC Sat 1 and SWC FWP 1) optimised more effectively than the UZ3 class, relatively permeable and free draining soils on permeable substrate with deep groundwater (SWC Sat 3 and SWC FWP 3). The ET depth parameter displays a clear optimisation below one meter and is the only parameter that is uniform in spatial distribution across the catchment. The horizontal hydraulic conductivities S4 - Kidderminster/ Chester Pebble Beds, S5 - Bridgnorth/ Kinnerton Formation, and S6 - Wildmoor/ Wilmslow Formation also show clear optimisation through the calibration process. The method of optimising the RMSE is therefore shown to result in the optimising and narrowing of the parameter limits in all cases apart from UZ class 3.

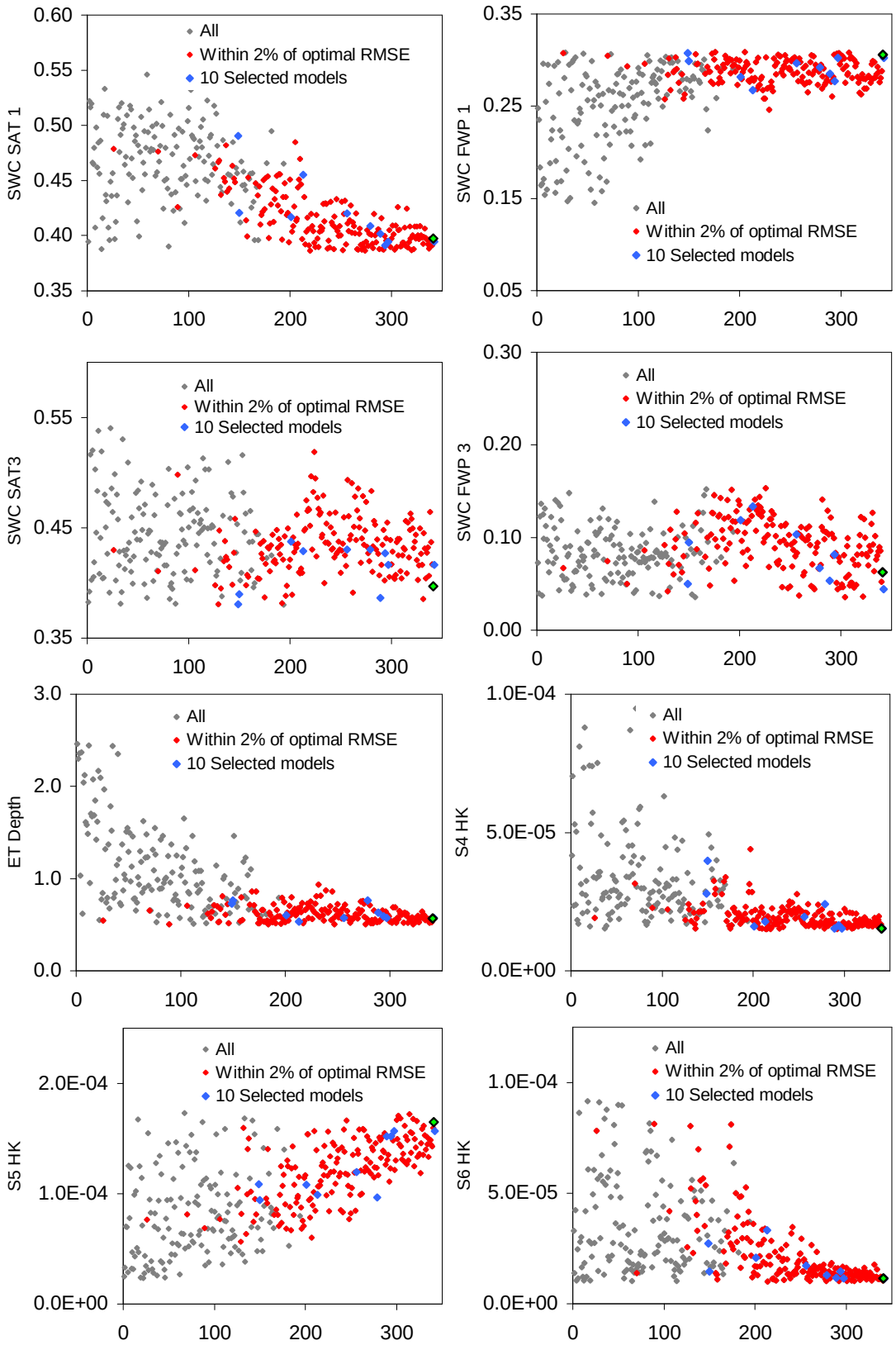


Figure 6.15. Optimisation of each parameter during automatic calibration of the distributed model. (Units for y axes: SWC Sat and SWC Fwp (fractions), horizontal hydraulic conductivity (m s^{-1}), ET Depth (m))

Table 6.13. Parameter values for ten sampled models within the context of minimum and maximum parameter ranges for the distributed model automatically calibrated at the basin outlet (unless stated, units are fractions)

ID	Run ID	SAT1	FWP1	SAT3	FWP3	ET (m)	S4 (m s^{-1})	S5 (m s^{-1})	S6 (m s^{-1})	RMSE flow Walcot m^3s^{-1}
1	342	0.395	0.302	0.417	0.044	0.571	1.538E-05	1.568E-04	1.152E-05	2.989
2	341	0.397	0.306	0.397	0.062	0.565	1.532E-05	1.651E-04	1.137E-05	2.989
3	294	0.391	0.277	0.427	0.081	0.591	1.652E-05	1.522E-04	1.468E-05	2.996
4	289	0.402	0.285	0.386	0.054	0.630	1.527E-05	1.520E-04	1.165E-05	3.000
5	201	0.417	0.282	0.438	0.119	0.601	1.607E-05	1.082E-04	2.084E-05	3.005
6	256	0.420	0.296	0.430	0.103	0.576	1.952E-05	1.196E-04	1.718E-05	3.009
7	279	0.409	0.292	0.430	0.067	0.761	2.413E-05	9.635E-05	1.289E-05	3.014
8	213	0.455	0.268	0.429	0.134	0.535	1.775E-05	9.899E-05	3.325E-05	3.021
9	150	0.421	0.299	0.390	0.095	0.763	3.981E-05	9.402E-05	1.446E-05	3.034
10	149	0.491	0.308	0.381	0.050	0.730	2.788E-05	1.083E-04	2.732E-05	3.048
Parameter ranges of the 10 sampled sets (above) within 2% of the optimal RMSE										
Minimum		0.391	0.2676	0.3806	0.04425	0.5346	1.53E-05	9.4E-05	1.14E-05	2.989
Maximum		0.491	0.3075	0.4376	0.13350	0.7633	3.98E-05	0.000165	3.33E-05	3.048
Parameter ranges of all 191 calibrated sets within 2% of the optimal RMSE										
Minimum		0.386	0.246	0.381	0.036	0.502	1.503E-05	5.628E-05	1.008E-05	2.989
Maximum		0.491	0.309	0.519	0.153	0.990	4.403E-05	1.719E-04	8.124E-05	3.048

SAT1 = Soil water content at saturated conditions and FC1 = Soil water content at field capacity for slowly permeable soils with prolonged seasonal water-logging over slowly permeable substrates.
SAT 3 = Soil water content at saturated conditions and FC3 = Soil water content at field capacity for relatively permeable and free draining soils on permeable substrate with deep groundwater
S4 HK = horizontal hydraulic conductivity of Kidderminster/Chester Pebble Beds, S5 HK= horizontal hydraulic conductivity of Bridgnorth/Kinnerton Formation, S6 HK= horizontal hydraulic conductivity of Wildmoor/Wilmslow Formation.

6.4.2.2. Calibrated model results and testing/validation

Output river flows and groundwater levels from the re-simulated suite of ten calibrated models are now presented. Figure 6.16 plots minimum and maximum calibrated bounds for the suite of ten sampled models at the same four gauging stations used in the manual calibration in Section 6.2, and calibration of the homogenous model in Chapter 5.

As in Chapter 5, the additional gauging stations of the Roden, Meese and Bailey Brook are used as a test and validation of the automatic calibration procedure to see how well the calibrated model (at the basin outlet) simulates river flow internally within the catchment. Figure 6.16 also includes the observed and manually calibrated model results to facilitate comparisons of model performance.

The upper and lower bounds of output river flow for the ten models from the automatic calibration at Walcot is shown to result in a narrow calibration band, where little

difference is seen between the output of the ten different models. This confirms that there can be a multitude of parameter sets from which almost equally statistically viable models can be derived. Notably, the SWC Sat parameters shown in Table 6.13 indicate a relatively wide calibrated parameter space for this parameter, which was also shown in the sensitivity analysis to be very important in determining the shape of output river flow and groundwater levels.

The independently calculated summary scores (Table 6.14) indicate that model performance at Walcot shows a slight improvement of 0.84, when compared to 0.80 in the manual calibration. The summary score is again in accordance with the other calculated statistics with only slight increases in ability as subsequently described, and indicates that expert elicitation (Refsgaard et al., 2006) although subjective, may be a useful tool especially when other more complex statistics may potentially be skewed by suspect input data (e.g. rainfall).

Upon inspection, the automatic calibration at Walcot simulates the winter peaks of November and December 2000 with a closer fit to the observed data (Figure 6.16). The baseflows appear similar for the automatic and manual calibration, with a slight decrease in simulated river flows seen for the automatic calibration. Aside of a slight improvement in RMSE, it is shown that at Walcot the NSE statistic ranges from 0.71 to 0.72 (the manual calibration NSE = 0.71) and the R ranges between 0.87 – 0.88 (compared to manual calibration of 0.87).

Table 6.15 summarises key statistics from the ten selected models based on the automatic calibration. For each site both the minimum and maximum values of mean river flow or groundwater level, total flow, Nash-Sutcliffe NSE, Correlation R, RMSE and RMSEP for the suite of ten calibrated models are shown.

Table 6.14. Results of summary score of river flows from automatic calibration at the basin outlet

Flow	**Walcot**	**Roden**	**Meese**	**Bailey Brook**
Baseflow	0.22	0.22	0.16	0.10
Peaks	0.16	0.21	0.22	0.23
Timing	0.23	0.23	0.20	0.17
MDF	0.23	0.16	0.20	0.12
Total	**0.84**	**0.82**	**0.78**	**0.62**
Mean score	**0.765**			

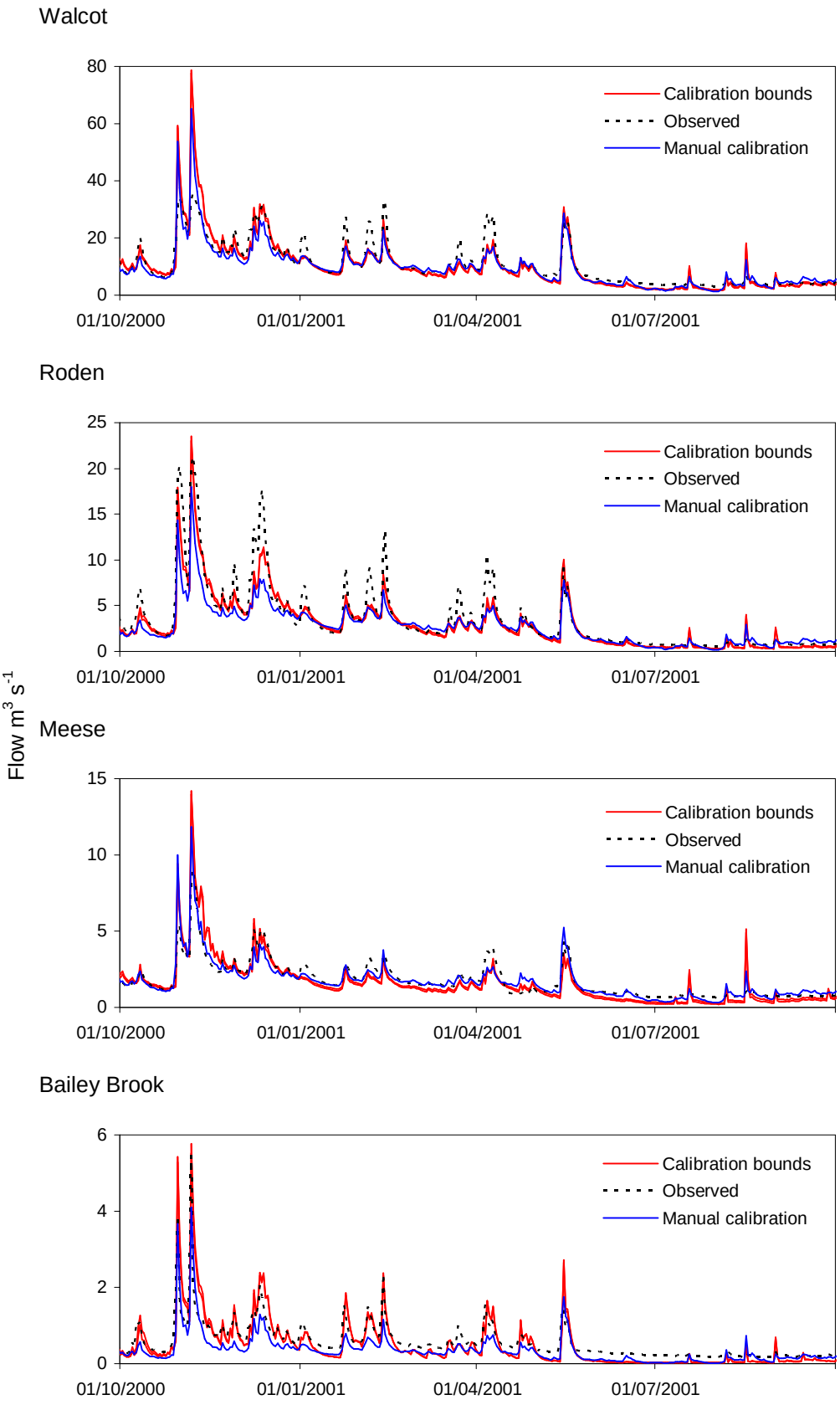


Figure 6.16. Observed and simulated calibration bounds at four river flow gauging stations for the distributed model calibrated at the basin outlet. The model is calibrated at Walcot and tested at the further three internal gauging stations

Table 6.15. Summary performance statistics for the distributed Tern model calibrated at the basin outlet

Name	Measure	Min - max MDF (m^3s^{-1}) or MGWL (m)	Range (m^3s^{-1}) or (m)	Total flow m^3s^{-1}	NSE	R	RMSE Flow (m^3s^{-1}) GW (m)	RMSEP
Walcot	Flow	6.90 – 7.11	0.21	8.30×10^8 - 8.55×10^8	0.71 – 0.72	0.87 – 0.88	2.99 – 3.05	0.406 - 0.414
Roden	Flow	1.79 – 1.89	0.10	2.26×10^8 - 2.39×10^8	0.81 – 0.83	0.91 – 0.92	0.99 – 1.03	0.463 - 0.481
Meese	Flow	1.07 – 1.14	0.07	1.35×10^8 - 1.44×10^8	0.52 – 0.53	0.85 – 0.86	0.56 – 0.57	0.463 - 0.471
Bailey Brook	Flow	0.30 – 0.32	0.02	3.82×10^7 - 4.02×10^7	0.10 – 0.13	0.82 – 0.83	0.31 – 0.31	0.795 - 0.795
Cherrington	GW *	54.50 – 55.60	1.1	NA	NA	0.19 – 0.43	2.17 – 3.25	NA
Edgmond	GW *	66.59 – 67.83	1.24	NA	NA	0.90 – 0.94	0.52 – 1.46	NA
Warren Farm	GW *+	77.59 – 80.16	2.57	NA	NA	0.59 – 0.72	2.46 – 4.79	NA
Gnosall	GW -	83.02 – 86.35	3.51	NA	NA	0.10 – 0.25	3.66 – 6.49	NA
Hawgreen	GW ×	64.20 – 67.69	3.49	NA	NA	0.49 – 0.59	2.98 – 6.35	NA
Heathlanes	GW ×	59.63 – 61.55	1.92	NA	NA	-0.07 – 0.20	1.82 – 3.71	NA
Longdon	GW ×+	51.76 – 52.20	0.44	NA	NA	0.80 – 0.83	1.52 – 1.95	NA

BK × – Bridgnorth Kinnerton

WW - – Wildmoor Wilmslow

KC * – Kidderminster Chester Pebble beds

D + –Sand/gravel drift deposits

Model performance at the internal un-calibrated gauging stations are also comparable to the manual calibration, also shown in Figure 6.16. A notable difference between the manual and automatic calibrations is the increase in the simulation of peak flows, especially in the Roden and in Bailey Brook for November and December 2000. Despite this, in the Roden tributary the automatic calibration fails to simulate the observed peaks during the first half of 2001, as did the manual calibration. This highlights that although slight improvements may be seen with the automatic calibration (the mean summary score, Table 6.14, showing a mean of 0.765 improved from 0.71 in the manual calibration) the fundamental larger issues with the model are still not improved or resolved. This again also raises the importance of assessing output plots compared to observed data, rather than relying solely on generated statistics.

Minimum and maximum output groundwater levels are displayed in Figure 6.17 for the same seven boreholes used in the manual calibration. In this automatic calibration, the

groundwater level data were not calibrated and as a result serve as a means to test the automatic calibration (as was also the case in Chapter 5). The plots compare the automatically calibrated outputs with manually calibrated outputs and the observed data. What is initially noted is that the calibration bounds are much wider for groundwater levels (when compared to flows that showed little variation in the output between the ten models). This is expected, as the boreholes were not used to calibrate the model and therefore the performance of groundwater levels (RMSE) has not been optimised in the procedure.

The range of performance statistics are shown in Table 6.15 for the seven boreholes. Compared to the gauging stations, the wider ranging statistics for each borehole are a reflection of the wider output bands. For example, the range of RMSE at Edgmond (shown in Figure 6.17 to have the narrowest calibration band) is from 0.52 to 1.46 m (Table 6.15). Comparatively, the RMSE at Gnosall (simulated with almost no ability, as shown in the output plot and supported with a summary score of 0.03), ranges from 3.66 to 6.49 m (Table 6.15).

Individual groundwater level plots in Figure 6.17 indicate that in only two boreholes is the performance of the model better as a result of the automatic calibration compared to the manual calibration. These sites are Gnosall and Warren Farm where the model shows very little ability at both sites regardless of the calibration method used. At all the other boreholes, the automatic calibration based on one objective function of RMSE of flow at the basin outlet results in poorer simulations of groundwater levels when compared to the manual calibration. Despite the boreholes being ‘un-calibrated’, in many cases the model is able to simulate the annual variability with fluctuations of draw down in the summer and re-charge in the winter, notably for Longdon and Edgmond sites where Table 6.15 supports this with good correlation ranges of 0.8 to 0.83 and 0.9 to 0.94, respectively.

The summary scores for groundwater levels simulated from this automatic calibration are provided in Table 6.16. The mean score is 0.39 compared to 0.43 for the manual calibration, a reflection that the model shows decreased ability at simulating groundwater levels when automatically calibrated using river flow at the basin outlet, as would also be expected.

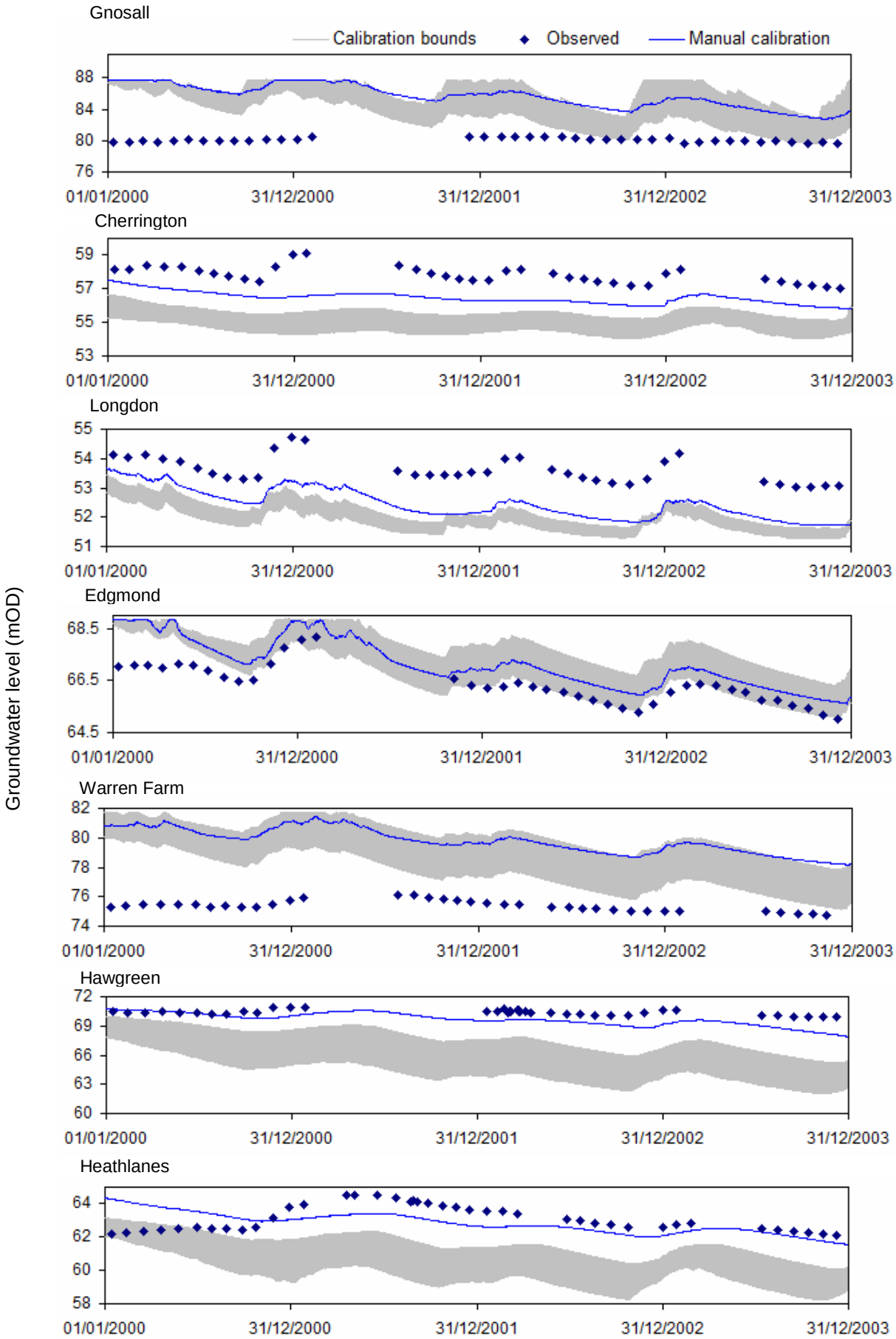


Figure 6.17. Results of calibration bounds derived by autocalibration of the distributed model of river flow at the basin outlet, for seven groundwater level boreholes that were not used in the calibration

Table 6.16. Summary score results for groundwater levels at seven boreholes for the automatic calibration at the basin outlet for the distributed model

	Cherrington	Edgmond	Warren	Gnosall	Hawgreen	Heathlanes	Longdon
General level	0.20	0.28	0.04	0.03	0.09	0.20	0.36
Overall shape	0.15	0.39	0.25	0.01	0.20	0.16	0.40
Total	**0.35**	**0.67**	**0.29**	**0.04**	**0.29**	**0.36**	**0.76**

Mean borehole summary score = 0.39

To summarise, the automatic calibration results for the distributed model that is automatically calibrated using output RMSE for river flow at the basin outlet indicates a model that is only slightly better at simulating river flows, but is unsurprisingly less adequate at simulating groundwater levels when compared to the manual calibration. The wide calibration bounds for simulated groundwater levels is expected to be a result of groundwater levels not being included in the optimisation process, and additionally as a result of the more complex nature of the distributed model.

6.4.3. Automatic calibration using a multi-criteria and multi-location approach

The same methodology outlined in Section 5.5.1 and Section 6.4.1 has been used for the second automatic calibration that again uses the objective functions of equally weighted RMSE of flow at four gauging stations, and the other objective function that equally weights the RMSE of groundwater levels at seven boreholes as shown in Figure 5.19. Convergence criteria were met after 555(+1 initial) simulations of the model that used different parameter sets in each run.

The mean optimal values of each RMSE objective function were 1.2215 m^3s^{-1} for flow, and 0.9910 m for groundwater levels. For comparison to the manual calibration and automatic calibration at Walcot, the optimal RMSE value for flow at Walcot that was achieved in this second automatic calibration was 3.056 m^3s^{-1} (RMSEP = 0.41) compared to the first autocal at the basin outlet, 2.989 m^3s^{-1} (RMSEP = 0.405), and manual calibration 3.07 m^3s^{-1} (RMSEP = 0.42).

6.4.3.1. Selection of calibrated models and parameter variation

To enable a fair comparison between models in the discussion of the thesis research questions in Chapter 7, the same method that was used in Chapter 5 has been used to select a threshold to determine a suite of calibrated models for this automatic calibration method.

The automatic calibration reports the individual parameter values, the RMSE of flows and the RMSE of groundwater levels for each of the 556 simulations carried out for this calibration method. Using the same 5% threshold from the optimal RMSE values that was described in Section 5.5.3.1 for the multi-objective automatic calibration of the homogenous model, this method allows models/parameter sets that have the lowest RMSE for both objective functions to be included as 'calibrated' models, and does not specifically highlight models/parameter sets that are better for simulating one or other of the objective functions (e.g. river flows or groundwater level RMSE).

Table 6.17 compares the thresholds for the different percentage changes allowed in the RMSE in classifying whether a model is considered as calibrated or not. As with the homogenous model in Section 5.5.3.1, it can be seen that different numbers of models/parameter sets are found for the different river flow and groundwater objective functions, with the number of models allotted within the flow objective function (534) greater than that of groundwater levels (201) for the 5% threshold.

Table 6.17. Selection of 5% from optimal RMSE threshold of calibrated models

Multi-criteria/multi-location distributed setup – 554 parameter sets tested
Optimal mean river flow RMSE (m^3s^{-1})= 1.2293
Optimal mean groundwater level RMSE (m) = 0.991

% increase from optimal		1%	2%	3%	4%	5%	10%
Increase allowed from optimal RMSE	Flow	0.0123	0.0246	0.0369	0.0492	**0.0615**	0.1229
	GWL	0.0099	0.0198	0.0297	0.0396	**0.0496**	0.0991
Threshold	Flow	1.2415	1.2538	1.2661	1.2784	**1.2907**	1.3522
	GWL	1.0009	1.0108	1.0207	1.0306	**1.0406**	1.0901
No. parameter sets within flow threshold		314	372	414	495	**534**	554
No. parameter sets within GWL threshold		34	82	121	161	**201**	327
No. parameter sets within both flow and GWL threshold		**34**	**82**	**121**	**161**	**201**	**327**
% of total runs		**6.1**	**14.8**	**21.8**	**29**	**36.3**	**59**

As already noted, and shown in bold in Table 6.17, the 5% threshold from optimal RMSE values for both flows and groundwater levels was selected. Although this threshold is again higher than for the automatic calibration at the basin outlet in the previous section (2%), the higher threshold accommodates for the inclusion of two objective functions, not one, that was characteristic of the previous automatic calibration described in Section 6.4.2. In total, 36.3% of the runs are considered as calibrated, statistically viable models. This is a large percentage but statistically only results in models that are within a 0.06 m^3s^{-1} mean change in RMSE for flow, and 0.05m mean change in RMSE for groundwater levels.

Table 6.18. Parameter values for ten sampled models within the context of minimum and maximum parameter ranges for the multi-objective automatically calibrated distributed model (unless stated, units are fractions)

Run	SAT1	FWP1	SAT3	FWP3	Depth of ET (m)	S4 (m s^{-1})	S5 (m s^{-1})	S6 (m s^{-1})	Mean Flow RMSE m^3s^{-1}	Mean GW RMSE m
315	0.392	0.250	0.383	0.073	0.552	9.646E-05	3.857E-05	5.996E-05	1.222	1.040
380	0.403	0.224	0.383	0.118	0.699	9.620E-05	4.273E-05	6.361E-05	1.235	1.035
406	0.423	0.187	0.383	0.143	0.684	9.620E-05	3.344E-05	6.967E-05	1.242	1.025
407	0.428	0.219	0.382	0.119	0.634	9.645E-05	3.734E-05	5.759E-05	1.235	1.040
420	0.412	0.187	0.380	0.110	0.702	9.522E-05	3.695E-05	7.208E-05	1.241	1.020
452	0.410	0.190	0.382	0.110	0.633	9.399E-05	3.337E-05	7.027E-05	1.236	1.034
484	0.397	0.156	0.383	0.131	0.538	9.670E-05	3.925E-05	6.360E-05	1.231	1.004
516	0.391	0.157	0.380	0.152	0.540	9.674E-05	3.760E-05	6.463E-05	1.231	0.995
548	0.397	0.186	0.379	0.151	0.555	9.617E-05	3.770E-05	6.576E-05	1.230	1.002
555	0.388	0.146	0.380	0.140	0.519	9.681E-05	3.905E-05	6.855E-05	1.229	0.991
Parameter ranges of sampled 10 parameter sets at 5% from RMSE										
Min	0.388	0.146	0.379	0.073	0.519	9.399E-05	3.337E-05	5.759E-05	1.222	0.991
Max	0.428	0.250	0.383	0.152	0.702	9.681E-05	4.273E-05	7.208E-05	1.242	1.040
Parameter ranges of all 201 calibrated sets at 5% from RMSE										
Min	0.386	0.141	0.378	0.053	0.503	9.201E-05	3.193E-05	5.353E-05	1.222	0.969
Max	0.441	0.268	0.393	0.153	0.793	9.718E-05	4.649E-05	7.693E-05	1.242	1.023

SAT1 = Soil water content at saturated conditions and FC1 = Soil water content at field capacity for slowly permeable soils with prolonged seasonal water-logging over slowly permeable substrates.
SAT 3 = Soil water content at saturated conditions and FC3 = Soil water content at field capacity for relatively permeable and free draining soils on permeable substrate with deep groundwater
S4 HK = horizontal hydraulic conductivity of Kidderminster/Chester Pebble Beds, S5 HK= horizontal hydraulic conductivity of Bridgnorth/Kinnerton Formation, S6 HK= horizontal hydraulic conductivity of Wildmoor/Wilmslow Formation.

As already noted in Chapter 5, the output of the automatic calibrations do not include many of the other statistics (such as NSE, R) that are being used as performance criteria within this thesis. For this reason, as well as to use the same methodology for the homogenous and distributed models, a further sample of ten of these models has been

undertaken and then re-simulated within MIKE SHE so as to generate the needed statistics. Once again, the sample is systematically generated by listing the different models in run order, and selecting every 28th model from the list after ensuring the inclusion of the optimal model and the two lowest groundwater and flow RMSEs within the suite of ten sampled models.

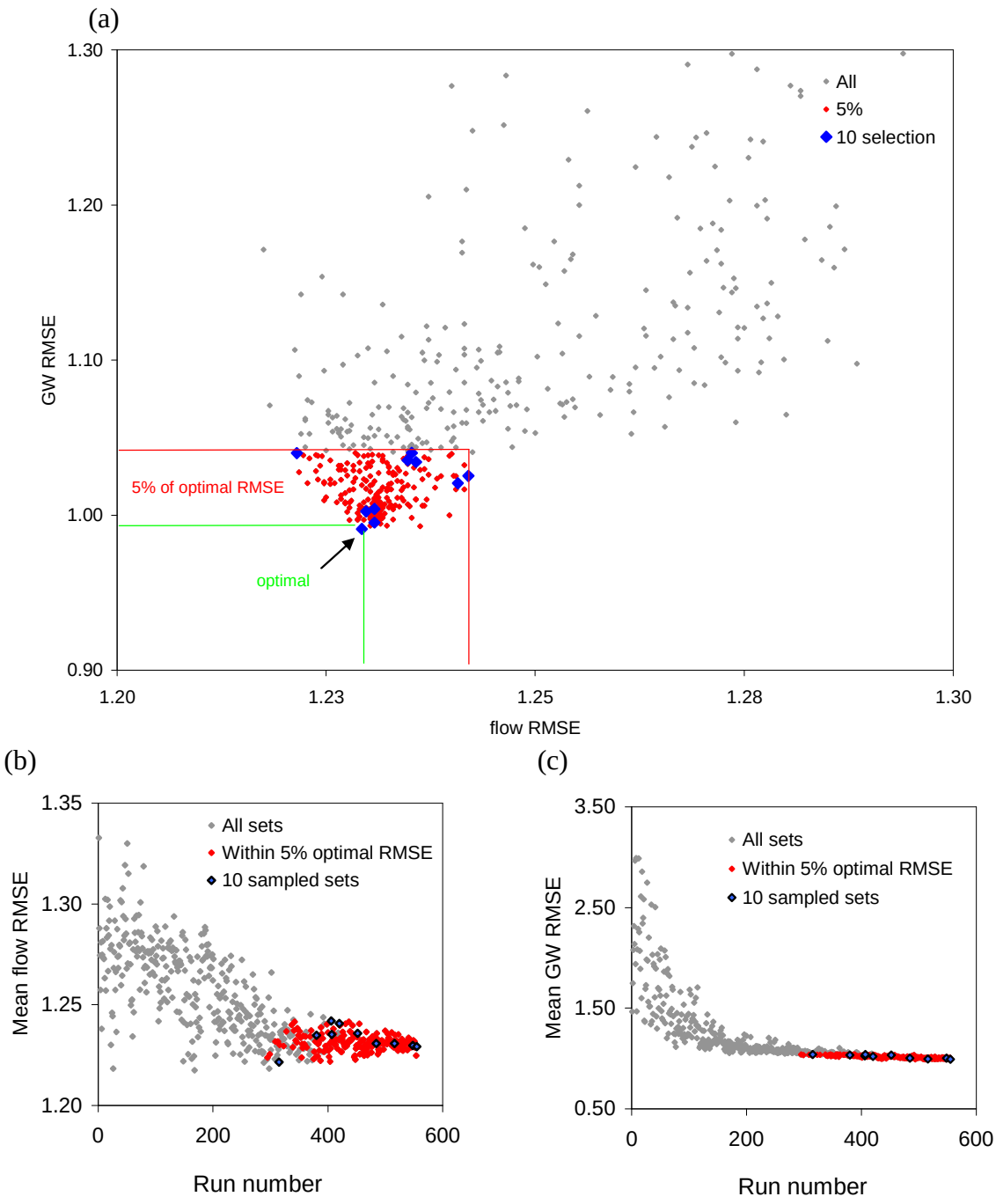


Figure 6.18. (a) Mean river flow and groundwater level RMSE identifying the optimal and sampled 'calibrated models' (b) Optimised mean river flow RMSE (m^3s^{-1}) (c) Optimised mean groundwater level RMSE (m)

The parameter values and parameter ranges, as well as mean river flow and groundwater RMSE are listed in Table 6.18 for the ten selected models. Figure 6.18a uses the output automatic calibration data to plot the mean river flow RMSE against the mean groundwater level RMSE. The plot shows the complete set of simulations as well as indicating the set of calibrated models (red), the sampled set of ten models (blue) as well as the mathematical optimum model. The threshold lines have been added to the plots to clarify the space of the plot from which the calibrated models are located. As shown, this calibrated region is located in the region close behind the optimal model.

The plots in Figure 6.18 b and c show the optimisation of the flow and groundwater level RMSE separately. Both plots show how the optimisation results in the improvement of the mean RMSE, with the narrow region toward the end of each optimisation including the range of calibrated and ten sampled models. The mean optimal RMSE values of 1.2215 m^3s^{-1} for flow, and 0.9910 m for groundwater levels are comparable values to the work undertaken by Madsen (2003) where mean RMSE values were 0.44 m^3s^{-1} for flow and 1.08 m for groundwater levels, for the Karup Catchment, Denmark.

Figure 6.19 presents the optimisation of the individual parameters that were subject to automatic calibration, showing also the selection of the suite of ten models, as well as all of the statistically calibrated parameter sets within the 5% from optimal RMSE. When compared to the previous automatic calibration optimised using flow RMSE at the basin outlet (Figure 6.15), the parameter optimisations in Figure 6.19 result in more efficiently optimised parameters, especially for the horizontal hydraulic conductivity parameters HK4, HK5 and HK6. This is likely due to the inclusion of the boreholes in the objective function as it was shown in the sensitivity analysis in Section 6.3 that the HK parameters were most influential in determining the shape of groundwater levels.

It is also notable from Figure 6.19 that some parameters are calibrated with more equifinality and range within the parameter space than others. For example, the SWC Fwp 1 and SWC Fwp 3 are considered calibrated with ranges of 0.146 to 0.250 and 0.073 to 0.152 respectively. Putting this into context, these parameter ranges (as well as the other ranges of parameters shown in Table 6.18) display that very different

parameter sets result in equally calibrated models. In reality, it is likely that the defined parameters are not realistically represented with one value in a model (as they

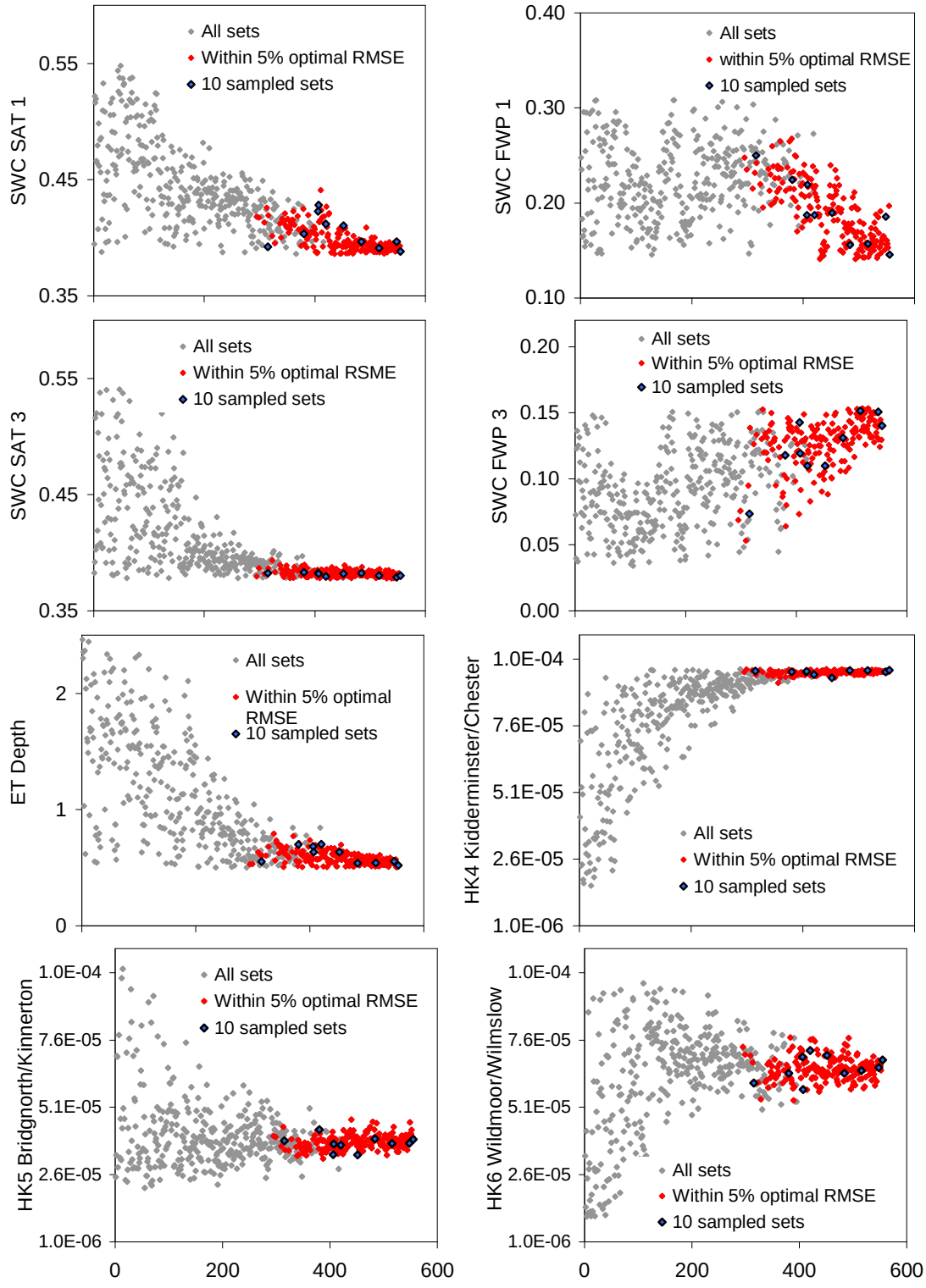


Figure 6.19. Parameter ranges tested within the automatic calibration and selected sample set of ten models. Units for y axes: SWC Sat and SWC Fwp (fractions), horizontal hydraulic conductivity (m s^{-1}), ET Depth (m)

inherently fluctuate spatially for the same parameter (i.e. SWC Sat is likely to vary from place to place and with different depths). The calibrated ranges shown therefore indicate that the model can work almost equally as well with different calibrations.

Having described the selection of calibrated models from the automatic calibration procedure to demonstrate equifinality, the output time series of river flows and groundwater levels, and statistics from the ten sampled models are presented and discussed.

6.4.3.2. Calibrated model results and testing/validation

The calibrated bounds (upper and lower simulated mean daily flow from the ten sampled models) for flows at the four river flow gauging stations used in the multi-objective automatic calibration of the distributed model, are shown in Figure 6.20. Additionally, the observed river flow data and manually calibrated results are also shown to enable comparisons to be made more easily.

Figure 6.20 indicates that the calibration bounds (shown in red) are very narrow for all four gauging stations. This means that the ten sampled models each with different calibrated parameter sets result in almost identical output river flows, again confirming the equifinality within the parameter sets. This result is comparable with the automatic calibration using only river flow RMSE at the basin outlet. The simulation statistics associated with the output river flow and groundwater levels are given in Table 6.19 where very similar results are seen when compared to the automatic calibration at the basin outlet. For example, the NSE of flow at Walcot in this calibration ranges from 0.70 to 0.71 compared to 0.71 to 0.72 for the distributed automatic calibration at the basin outlet. The statistics are also similar for the other internal gauging stations between the two automatic calibrations, which is interesting as the observed data at these three sites were included in the objective function used in the calibration for this method, whereas they were not used in the previous method.

Figure 6.20 also shows that this multi-objective automatic calibration results in a better fit with the observed river flow data compared to the manual calibration. Of all the gauging stations, the results in the Meese are very similar for the two methods of calibration, with the NSE increasing to 0.53 to 0.54 for the automatic calibration

compared to 0.50 for the manual calibration. This is also reflected in the similar MDF simulated at 1.19 to 1.21 in the automatic calibration and 1.21 in the manual calibration.

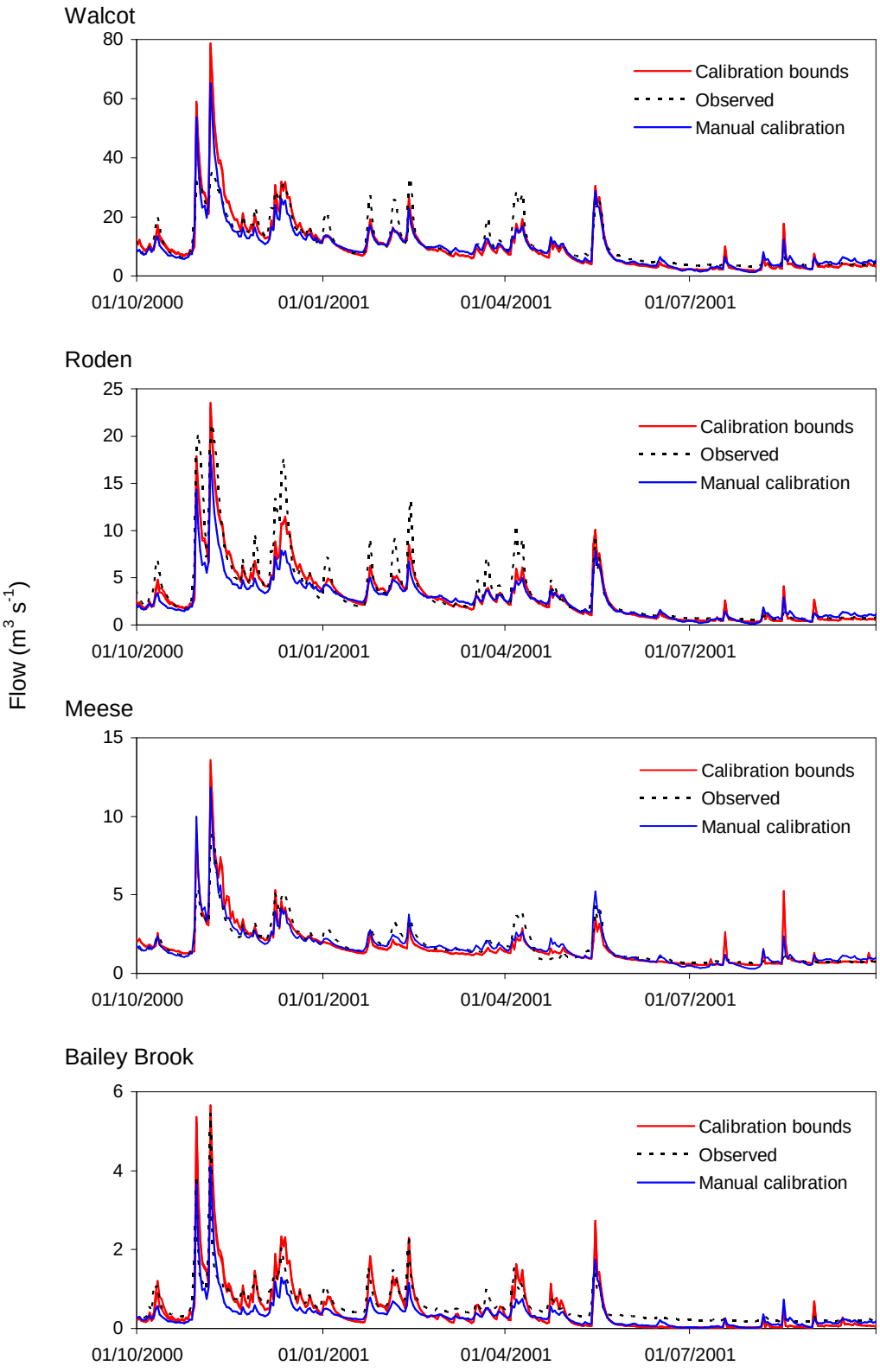


Figure 6.20. Observed and simulated daily river flow showing the calibration bounds (minimum and maximum daily flow) for the ten sampled models automatically calibrated with the multi-objective approach. Plots shown for the four gauging stations used in the calibration

The automatic calibration appears to result in most improved simulations in the Roden and Bailey Brook. Inspection of the plots in Figure 6.20 indicates that the peaks at Bailey Brook are simulated with much better ability, with a significant improvement especially in February 2001. Despite these improvements, the simulation of the baseflow during the summer (June-August 2001) is still considerably under-simulated by the automatic calibration. In the Roden, the timing and magnitude of flows in November and December 2000 are considerably improved; however there is little improvement in simulating the February 2001 peak flows.

As suggested previously, this indicates that although the automatic calibration can result in improved simulations in parts of the hydrograph, some of the more fundamental errors (such as potentially suspect or erroneous input rainfall data) in the simulation are not changed by any of the methods of calibration assessed.

Table 6.19 suggests that although performance is slightly improved for the multi-objective automatic calibration (such as in the Roden with a 0.98 to 1.01 m^3s^{-1} range in RMSE (RMSEP = 0.458 – 0.472) compared to the manual calibration (Table 6.3) of 1.05 m^3s^{-1} RMSE (RMSEP = 0.49), the statistical improvement is not large. The summary score values for the same calibration are presented in Table 6.20 where a larger improvement is seen in the mean score 0.81, compared to the manual calibration mean summary score, 0.71 (Table 6.2). Additionally, the separate scores for each of the tributaries are scored higher than for the manual calibration. The reason the summary score is higher may be due to it being a calculation that considers separately, then sums the criteria from four parts of the simulation (Table 5.2). Many of the more quantitative statistics are often biased by anomalies in the simulation (i.e. from one over-simulated event that could be caused by suspect precipitation data), whereas the summary score is a more balanced measure that considers the whole shape of the simulation.

Table 6.19. Statistical performance results from the calibration of the multi-objective distributed Tern model, for river flows and groundwater levels used in the objective functions/calibration. Results show the upper and lower (max and min) bounds for the range of the ten sampled models

Name	Measure	Min/max MDF m^3s^{-1}) or MGWL (m)	Range (m^3s^{-1}) or (m)	Total flow m^3s^{-1}	Nash-Sutcliffe NSE	Correlation R	RMSE Flow (m^3s^{-1}) GW (m)	RMSEP
Walcot	Flow	6.84 - 6.98	0.14	$8.21x10^8$ - $8.38x10^8$	0.70 – 0.71	0.87 - 0.87	3.04 – 3.09	0.412 - 0.419
Roden	Flow	1.90 - 1.94	0.04	$2.40x10^8$ - $2.45x10^8$	0.82 – 0.83	0.92 - 0.92	0.98 – 1.01	0.458 - 0.472
Meese	Flow	1.19 - 1.21	0.02	$1.50x10^8$ - $1.53x10^8$	0.53 – 0.54	0.81 - 0.82	0.56 – 0.56	0.463 - 0.463
Bailey Brook	Flow	0.30 - 0.31	0.01	$3.74x10^7$ - $3.92x10^7$	0.12 – 0.13	0.82 - 0.83	0.31 – 0.31	0.795 - 0.795
Cherrington	GW in KC	56.56 - 56.77	0.21	NA	NA	0.38 - 0.49	1.10 – 1.28	NA
Edgmond	GW in KC	66.12 - 66.26	0.14	NA	NA	0.86 - 0.87	0.38 – 0.43	NA
Warren Farm	GW in KC with sand/gravel D	76.49 - 76.92	0.43	NA	NA	0.51 - 0.52	1.44 – 1.76	NA
Gnosall	GW in WW	79.83 - 80.64	0.81	NA	NA	0.09 - 0.11	1.69 – 1.91	NA
Hawgreen	GW in BK	70.41 - 70.78	0.37	NA	NA	0.42 -0.61	0.34 – 0.48	NA
Heathlanes	GW in BK	62.89 - 63.26	0.37	NA	NA	-0.02 - 0.12	0.78 – 0.87	NA
Longdon	GW in BK with sand/gravel D	52.69 - 52.88	0.19	NA	NA	0.88 - 0.91	0.92 – 1.09	NA

BK – Bridgnorth Kinnerton, WW – Wildmoor Wilmslow, KC – Kidderminster Chester Pebble beds, D – Drift deposits

Table 6.20. Summary score for four river flow gauging stations in the multi-objective automatic calibration of the distributed model

Flow	Walcot	Roden	Meese	Bailey Brook
Baseflow	0.21	0.23	0.20	0.10
Peaks	0.19	0.21	0.20	0.20
Timing	0.23	0.22	0.22	0.18
MDF	0.22	0.20	0.24	0.19
Total	**0.85**	**0.86**	**0.86**	**0.67**
Mean score	**0.81**			

Overall, it is suggested that the multi-objective method of automatic calibration results in output river flow simulations that are considered fair. Although there are still discrepancies in some parts of the simulations, there are notable improvements in other parts which are of a similar standard to those discussed in the first automatic calibration using flow RMSE at the basin outlet. In order to better compare the methods of calibration, performance of the groundwater level simulations are now discussed.

Figure 6.21 presents calibrated bounds of the ten sampled models for the seven groundwater level boreholes used as output measures in the groundwater objective function within this calibration. The observed data and manual calibration are also shown as a means of comparison. As with the river flow simulation, the calibrated bounds of upper and lower daily groundwater levels result in a narrow band, highlighting that the ten sampled model result in very similar output levels despite having different parameter calibrations. For example, the statistics in Table 6.19 suggest mean daily range of mean groundwater levels are 0.14m at Edgmond (narrow range) compared to 0.81m at Gnosall (larger range). These ranges compare to much larger variation in the automatic calibration at the basin outlet (Table 6.15) where ranges of 1.24m at Edgmond, and 3.51m at Gnosall are seen. The reduction in the calibrated range in the multi-objective autocal is a result of the inclusion of the observed groundwater levels in the objective function. Therefore these groundwater levels have been included in the calibration whereas previously they were uncalibrated and used as a test or validation of the automatic calibration.

In general the plots in Figure 6.21 suggest a much improved simulation of groundwater levels as a result of this automatic calibration method. When compared to Figure 6.17 (for the single objective autocal at the basin outlet), a closer fit with the observed data is seen. This is supported with substantial improvements in the RMSE of groundwater levels such as at Gnosall with a range here from 1.69 to 191m compared to Table 6.15 where the RMSE range was 3.66 to 6.49m. The ranges at all the other boreholes also show comparative and similar improvements to the RMSE as a result of the multi-objective calibration.

In comparison to the manual calibration, Figure 6.21 suggests that the automatic calibration again results in substantial improvements to the models ability to simulate groundwater levels, in all boreholes, again this is due to these boreholes being included within the objective function of the automatic calibration. Despite the improvements, it is also noted that some of the general levels and shapes of the simulations representing the seasonal and annual variability are poor. This is reflected by the poor correlation, R, at Gnosall from 0.09 to 0.11 and at Heathlanes -0.02 to 0.12. At these two boreholes, there is almost no ability at representing the seasonal variations of levels, although the general level is improved from the manual calibration. At Cherrington the simulated levels show almost no seasonal or annual variability, apart from recharge at start of 2003. This is also reflected in the low R of 0.38 to 0.49 (Table 6.19).

The summary score for groundwater levels from the multi-objective automatic calibration are shown in Table 6.21. The mean score is 0.58 which although not as high as the corresponding score for river flows from this same calibration, (0.81, Table 6.20), is a considerable increase compared to the manual calibration mean score of 0.43 (Table 6.4), and the single objective autocalibration summary score of 0.39 (Table 6.16). The summary score values are higher when groundwater levels are specifically included in the automatic calibration. However, when the automatic calibration does not include levels in the optimisation procedure the mean summary score is reduced (and to a lower level than the manual calibration).

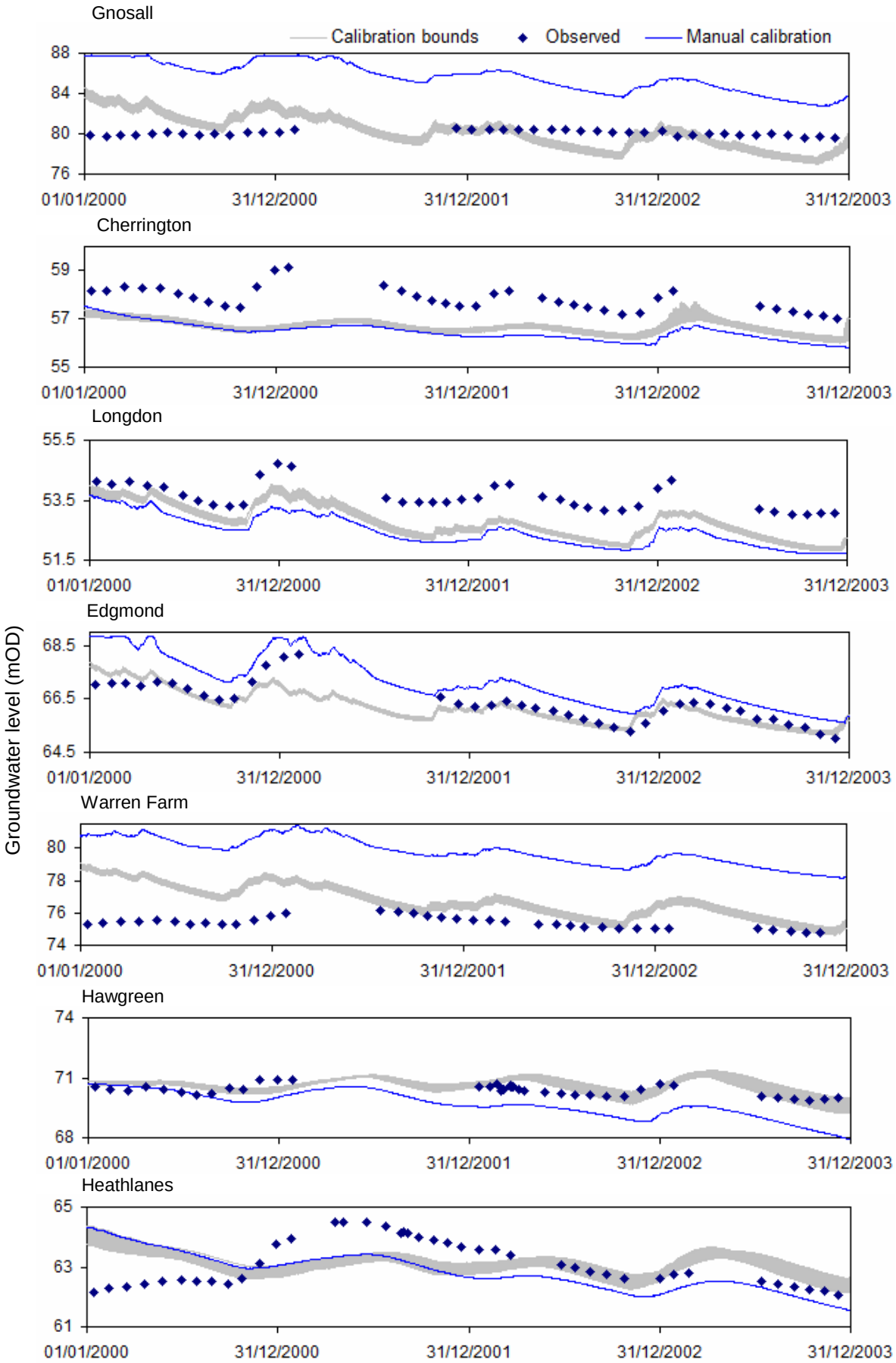


Figure 6.21. Results of the multi-objective automatic calibration of the distributed model. Observed and simulated groundwater levels from the range of ten sampled models are shown within the calibration bounds

Using the same method as in Chapter 5 for the homogenous model, in order to assess the robustness of the automatic calibration results, and how good simulations are at other non-calibrated sites, a testing/validation stage now uses the same additional gauging stations and groundwater boreholes that were used to test and validate the manual calibration in Section 6.2.3.

Simulations of river flow at the 'uncalibrated' gauging stations are shown in Figure 6.22 which plots the minimum and maximum range of mean daily flow for the suite of ten models as well as the observed data, and results of simulations from the manual calibration. When compared to the manual calibration and observed data there does not appear to be a consistent pattern in improved or reduced ability. At Eaton-on-Tern the model appears to simulate the peak flows with better accuracy, with the NSE increasing from 0.58 in the manual calibration (Table 6.6) to a range of 0.63 to 0.65 for the suite of ten automatically calibrated models, as shown in Table 6.22. Despite this, the model appears to over-simulate the mean daily flow, both at Eaton as well as at Ternhill (expected as these sites are both located on the main Tern channel). The correlation between simulated and observed data at these two sites are between 0.88 and 0.89 and 0.82 and 0.83 respectively, which are comparable to the correlation seen in Table 6.6 for the manual calibration.

Table 6.21. Summary score at seven groundwater level boreholes in the multi-objective automatic calibration of the distributed model

	Cherrington	**Edgmond**	**Warren**	**Gnosall**	**Hawgreen**	**Heathlanes**	**Longdon**
General level	0.35	0.4	0.3	0.2	0.38	0.35	0.4
Overall shape	0.2	0.39	0.2	0.1	0.2	0.22	0.4
Total	**0.55**	**0.79**	**0.5**	**0.3**	**0.58**	**0.57**	**0.8**

Mean borehole summary score = 0.58

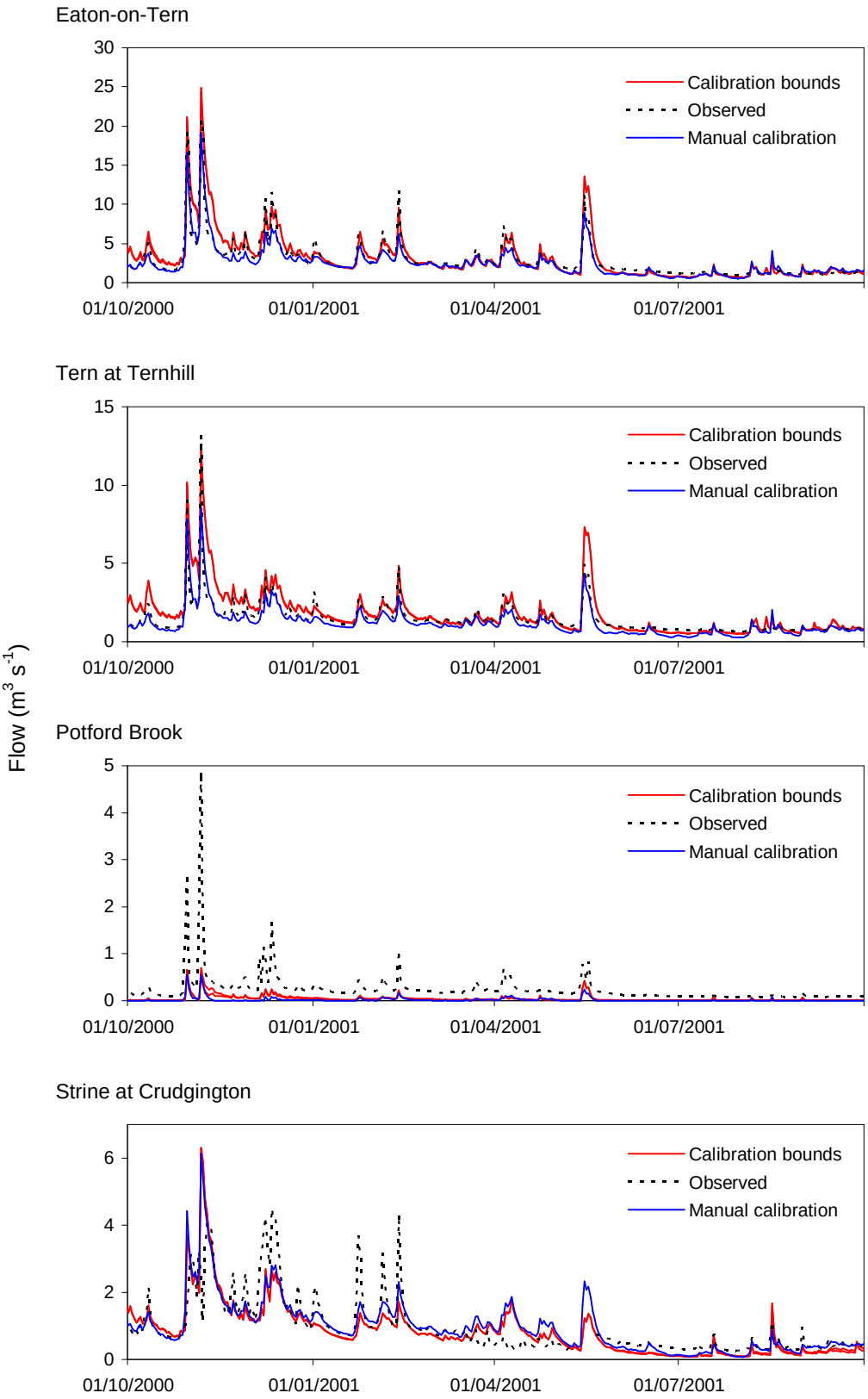


Figure 6.22. Testing of the multi-objective automatic calibration. Plots for four 'non-calibrated' gauging stations show observed and simulated river flows (calibrated bounds derived from the min & max daily river flow from ten sampled models)

Table 6.22. Statistical performance results from the testing/validation of the multi-objective distributed Tern model, for flows and groundwater levels not assessed within the calibration methodology. Results show the upper and lower (max and min) bounds for the range of the ten sampled models

Name	Measure	Min/max MDF (m^3s^{-1}) or MGWL (m)	Range (m^3s^{-1}) or (m)	Total flow m^3s^{-1}	Nash-Sutcliffe NSE	Correlation R	RMSE Flow (m^3s^{-1}) GW (m)	RMSEP
Eaton on Tern	Flow	2.18 - 2.23	0.05	$1.98x10^8$ - $2.24x10^8$	0.63 – 0.65	0.88 – 0.89	0.96 – 0.98	0.492 - 0.503
Tern at Ternhill	Flow	1.26 - 1.29	0.03	$9.24x10^7$ - $1.07x10^8$	0.11 – 0.19	0.82 – 0.83	0.62 – 0.65	0.590 - 0.619
Potford Brook	Flow	0.02 - 0.03	0.01	$1.30x10^6$ - $2.59x10^6$	-0.16 – -0.13	0.68 – 0.69	0.22 – 0.23	1.375 - 1.438
Strine	Flow	0.59 - 0.61	0.02	$7.52x10^7$ - $8.27x10^7$	0.57 – 0.58	0.79 – 0.80	0.38 – 0.39	0.528 - 0.542
Cotton Farm	GW in BK with silty clay D	64.97 - 65.07	0.10	NA	NA	0.88 – 0.89	1.61 – 1.73	NA
Radmoor	GW in BK with diamicton D	67.59 - 67.99	0.40	NA	NA	0.77 – 0.89	2.00 – 2.5	NA
Rowton	GW in BK silty clay D	61.27 - 61.61	0.34	NA	NA	0.26 – 0.33	1.02 – 1.23	NA
Heathcote	GW in BK with sand/ gravel D	72.35 - 72.60	0.25	NA	NA	0.81 – 0.84	1.37 – 1.47	NA
Woodlands Farm	GW in KC	67.45 - 67.71	0.26	NA	NA	0.58 – 0.59	2.67 – 2.93	NA
Twinney Lane	GW in KC	60.46 - 60.76	0.30	NA	NA	0.12 – 0.19	0.81 - 0.94	NA
Coley Farm	GW in WW with sand/ gravel D	73.17 - 73.17	0.00	NA	NA	0.73 – 0.74	2.12 – 2.12	NA

BK – Bridgnorth Kinnerton, WW – Wildmoor Wilmslow, KC – Kidderminster Chester Pebble beds, D – Drift deposits

The performance of the model in the Strine appears to show the opposite pattern, with a decrease seen in the simulated river flows in Figure 6.22 when compared to the manual calibration. Despite this, the performance statistics are not notably different when compared to the manual calibration in Table 6.6. The performance in Potford Brook tributary again shows little difference to the manual calibration, indicating that performance at uncalibrated gauging stations is not improved with the automatic calibration methodology.

Figure 6.23 compares the output groundwater levels to the observed data at the seven boreholes that are used for model testing (uncalibrated sites). The manually calibrated groundwater levels are also plotted for a comparison (from Figure 6.10) to assess if the automatic calibration procedure resulted in improved groundwater levels.

Given that the results of simulated groundwater levels were so improved in the multi-objective automatic calibration in Figure 6.21, the results in Figure 6.23 for the uncalibrated sites are important to note. In almost all cases, the model appears reduced in ability to simulate uncalibrated groundwater levels when compared to the manual calibration also shown within the plots. Only at Rowton could the model be described as simulating the levels more closely to the observed data than the manual calibration. Despite this, the model shows no ability at simulating the seasonal or annual fluctuation at this borehole and so it cannot be said that the simulation is good at this site.

Despite some of the fair correlation statistics shown in Table 6.22 (i.e. for Radmoor, Rowton and Heathcote), the output plots in Figure 6.23 are clear in demonstrating the lack of ability in the model in simulating groundwater levels outside of the calibration.

6.4.4. Review of the automatic calibration of the distributed model

Referring back to the aims outlined at the start of Section 6.4, the primary objective was to test if automatic calibration methodologies resulted in improved simulations of output flow and groundwater levels at a variety of locations within the catchment.

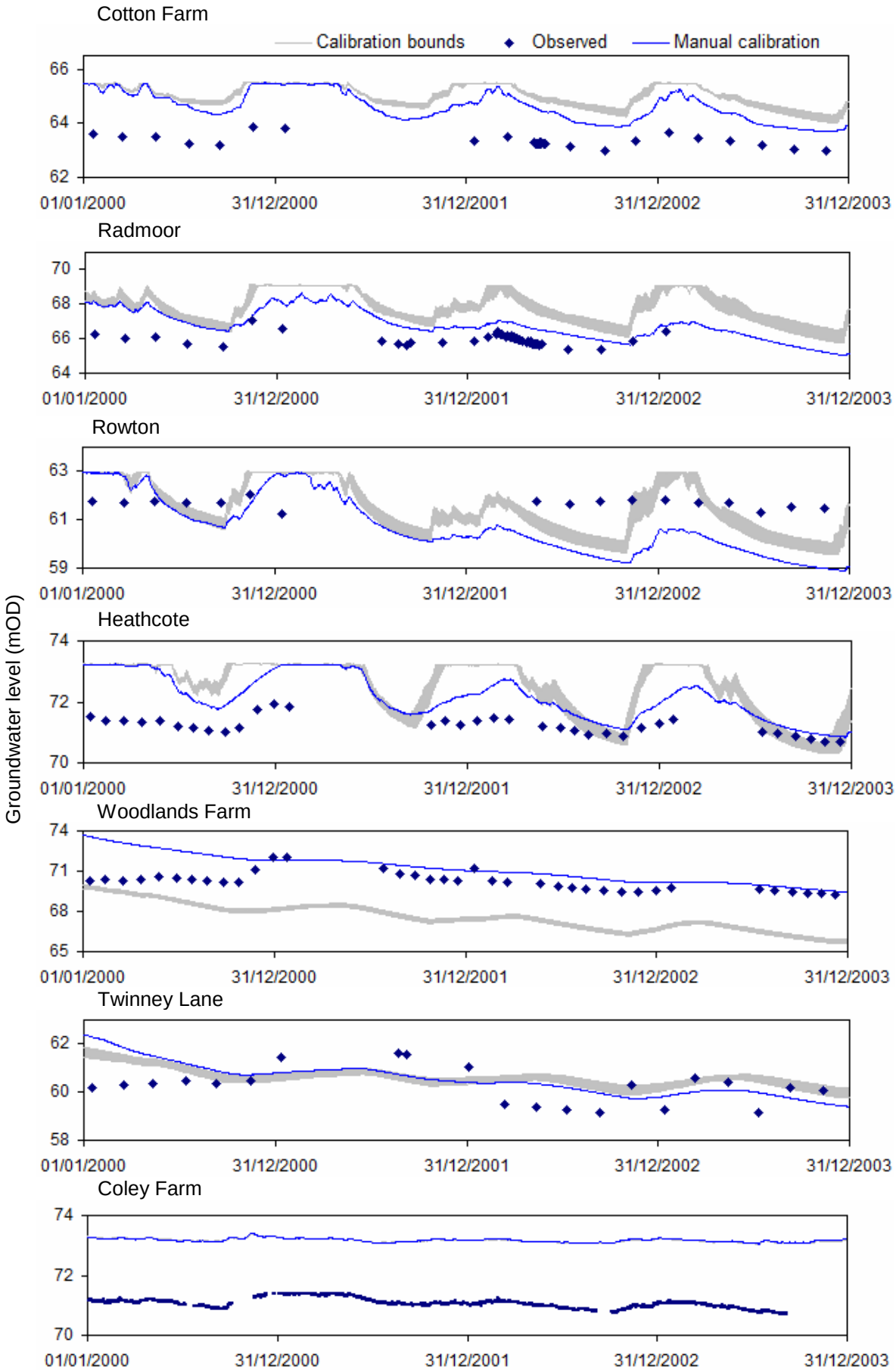


Figure 6.23. Testing/validation of the multi-objective automatic calibration. Plots show observed and simulated groundwater levels at seven 'non-calibrated' boreholes (calibrated bounds derived from the simulated min and max groundwater levels from the sampled ten calibrated models)

As with the method undertaken in Chapter 5 for the homogenous model, the calibrated minimum and maximum bounds for each set of ten models for both automatic calibrations have been directly compared to the manually calibrated model. For river flow simulations it has been shown that, in general, all three methods of model calibration result in similar performance statistics. For both the NSE and the R, the statistics varied little between the three methods. Despite this, the summary score calculation clearly highlights that the multi-objective automatic calibration method resulted in the best performance, scored at 0.81 compared to the manual calibration, 0.71 and single objective autocal, 0.77. It is noted that, as expected, the best simulation of the model at the basin outlet is shown by the model that is automatically calibrated to optimise the RMSE at Walcot. Comparatively, for groundwater level simulations, the mean RMSE for the same method of automatic calibration is large, at 4.00m.

These results draw attention to the benefit of having shown a range of calibration methods, as the mean RMSE for groundwater levels is optimised, at best to 0.9910m for the multi-objective automatic calibration. In addition, the other performance statistics have also indicated the better performance of the multi-objective autocal models at simulating groundwater levels. Both the quantitative statistics as well as the summary scores have therefore suggested that the automatic calibration methods result in a better calibrated model that is more balanced according to mathematical criteria in the objective functions. To a large extent this is to be expected as the distributed model is spatially much more complex and has many more model parameters that in the manual calibration were described as difficult to calibrate. However, the results of the testing and validation are not conclusive in suggesting that the automatic calibration methods are the best.

The second aim was to compare the complexity of the automatic calibration methods, and whether if including more rigorous multi-objective functions (in the second test) the ability of the model simulations of river flow and groundwater levels are improved. From the discussion above, results indicate that there is an apparent trade off in simulation ability with the two different automatic calibrations, where it is possible to calibrate a model in a way (from the autocal of RMSE at the basin outlet) that maximises the performance of river flows. However, the qualitatively more balanced method where both river flows and groundwater levels are simulated fairly well is found from the multi-objective automatic calibration.

Lastly, the automatic calibrations sought to assess and quantify some of the parameter equifinality within hydrological modelling. As for the homogenous model in Chapter 5, this has been undertaken using defined thresholds – that if a parameter set resulted in a model within a prescribed percentage of the optimal RMSE (2% for the single objective autocal and 5% for the multi-objective autocal), then it was considered as a calibrated model. Ten models each with different parameter sets were sub-sampled, plotted and compared for each automatic calibration method. The plotted flow and groundwater levels, especially for the multi-objective autocal resulted in very similar plots with narrow calibrated ranges. Despite the similar output flow and groundwater levels for all ten models, it was shown that the calibrated parameter sets were comprised of different parameter values, especially for the soil water content parameters in the unsaturated zone.

6.5. Summary

This chapter described the calibration and sensitivity analysis of the distributed Tern model set up that was described in Chapter 4. The same performance criteria that were described and used in Chapter 5 for the homogenous model were again used in this chapter for the distributed model to ensure a fair comparison of the two models in the following chapter.

The manual calibration of the distributed model was shown to be fair but not good, for both river flows and groundwater levels. This manually calibrated model was then used in an automatic sensitivity analysis within the Autocal component of MIKE ZERO. The sensitivity analysis that was undertaken for 61 model parameters resulted in the selection of seven model parameters that were then taken forward and used similar automatic calibration methods that were described in Chapter 5.

The multi-objective method was qualitatively the most balanced but does not statistically perform the best, especially for river flows. Therefore, it is difficult to say that one method of calibration is better than the rest as this chapter has also indicated that although all three methods of calibration result in generally fair simulations, the testing and validation of the calibrated set of models at other internal river flow and groundwater level boreholes resulted in poorer performance for all methods of calibration assessed.

Chapter 7

Issues of hydrological model structure, calibration and performance

7.1. Introduction

Having presented the results of manual calibrations, sensitivity analyses and automatic calibrations of both the Tern homogenous and Tern distributed models in Chapters 5 and 6 respectively; this chapter draws together the results from these two chapters. It discusses the issues associated with each of the research questions outlined in Chapter 1 in order to assess the main aim of the thesis of whether it is possible to select a best modelling protocol from those assessed. The research questions addressed in this chapter are:

1. What effects do different spatial representations in the model setup have on model outputs?

2. How do different methods of model calibration affect the performance of simulated river flows and groundwater levels?

3. How different are output river flows and groundwater levels as a result of parameter equifinality

4. To what extent do different performance statistics suggest different abilities of models?

5. How do measures of performance internally within the catchment, as well river flow at the basin outlet suggest different abilities of models?

The focus of the discussion in this chapter is given to the comparison of different spatial complexities of models and calibration methods. Therefore only the optimal models derived in the previous chapters are taken forward and compared in the analysis (i.e. one model from each calibration method and model complexity) as indicated in Table 7.1. In the case of the automatically calibrated models, the optimal model from each protocol is taken as the model with the lowest resulting RMSE (the criteria used for assessment in the automatic calibration process).

Table 7.1. The six calibrated models highlighting the spatial complexity and method of calibration for each.

Set up No.	Spatial complexity	Method of calibration	Model selected
1	Homogenous	Manual	Subjective optimal
2	Homogenous	Automatic using basin outlet	Statistically optimal
3	Homogenous	Automatic using multi-objective measures	Statistically optimal
4	Distributed	Manual	Subjective optimal
5	Distributed	Automatic using basin outlet	Statistically optimal
6	Distributed	Automatic using multi-objective measures	Statistically optimal

Section 7.2 presents summary result figures and tables that compare simulated river flows and groundwater levels from the six models with different protocols. The summary performance statistics for both calibration and validation sites are presented for both river flows and groundwater levels. The results are then discussed with reference to each of the research questions in Section 7.3. For each question the methodology that has been used in this thesis to explore the issue is reviewed, as well a discussion and evaluation given. Section 7.4 provides a brief summary and answers the main research aim of the thesis.

7.2. Compilation of results

Having provided independent figures and tables from the results of the different calibrations and model complexities in chapters 5 and 6, the figures and tables in this chapter summarise the uncertainty and differences in performance between the six different modelling protocols.

To demonstrate the range in calibrated uncertainty, Figures 7.1 (river flows) and 7.2 (groundwater levels) plot the daily minimum and maximum calibrated ranges for a number of gauging stations and boreholes for river flows and groundwater levels, respectively. Figure 7.1 indicates that at the basin outlet (Walcot), uncertainty in the

output river flows between the six models is quite small. For example, the range in the simulated mean daily flow of the six models at the basin outlet resulted in a 10.4% variation in simulated mean daily flow as a proportion of the mean daily flow of the six models. At internal gauging stations however (and especially at the smallest, Bailey Brook) calibrated river flows differ more substantially with a 43.6% variation in simulated mean daily flow as a proportion of the mean daily flow of the six models at Bailey Brook. This result indicates the heterogeneous response of the hydrological models within the catchment – and that what is seen at the basin outlet is not indicative of internal model performance, thus demonstrating and confirming the need of multi-site model assessment (Refsgaard, 1997).

Similar results are shown in Figure 7.2 which quantifies the uncertainty in calibrated groundwater levels. Different calibrated ranges for the six models are evident at different boreholes. At Longdon and Edgmond, the six models that the uncertainty bounds are comprised of show fair ability in simulating both the general level and shape compared to the observed data (with the mean range/difference on groundwater levels being 2.29m and 1.62m for the six models, for Longdon and Edgmond respectively). Comparatively, at Gnosall a range of 6.56m in the mean groundwater level is seen between the six different calibrated models.

For the purpose of showing how each of the summary score statistics were compiled, comparative scores are shown in Tables 7.2 for river flows for both calibration and validation sites and 7.3 for groundwater levels, for both calibration and validation sites. For each modelling protocol the individual components of the score are given as well as the total scores. In order to compare each of the modelling protocols further, river flow and groundwater level regimes are presented in Figures 7.3 and 7.4, respectively. These figures are discussed further in Section 7.3.

Table 7.4 compares the different performance statistics for all six models, for river flows. Table 7.5 shows the performance statistics for all six models for groundwater levels. Both tables distinguish between river flows and groundwater levels used for either the calibration or validation stages. Both the calibration and validation site data are shown in the performance tables to enable discussion about model performance when the models have each been calibrated differently and set up with different spatial representations (the different protocols). It is expected that performance for the

validation sites should be better for the more robust calibration methods that use multi-site river flow and groundwater level data, rather than the manual calibrations in which there is no mathematical optimisation.

The statistics in the performance tables are colour coded according to performance criteria shown in Table 7.6 in order to facilitate the interpretation. The colour codes seek to grade the model performance according to criteria suggested from other studies using the MIKE SHE modelling code. Henriksen et al., (2008) used a star system to assess statistical model performance, whilst Andersen et al., (2001) used a numbered rating system. Table 7.6 show the thresholds of classification for each statistic, with the Nash-Sutcliffe NSE and RMSE classifications set the same as Henriksen et al., (2008), and the percentage error performance classification the same as Andersen et al., (2001). The colour coded performances are also associated with numbered ratings in Table 7.6. Performance that is classified as excellent is associated with a score of 5, whereas performance that is rated as poor is given a scoring of 1.

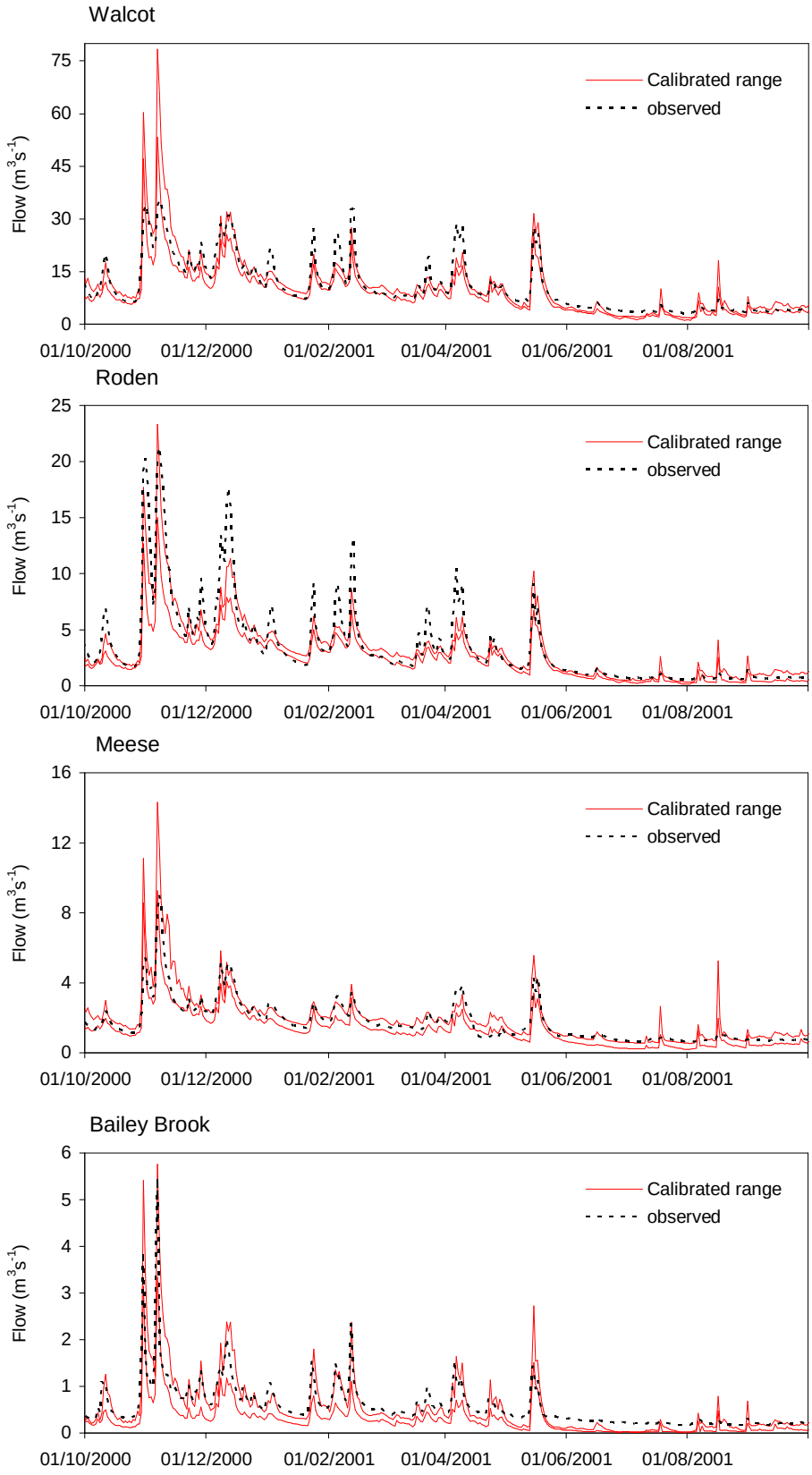


Figure 7.1. Uncertainty bounds in calibrated river flow at four gauging stations derived from six calibrated models of different modelling protocols

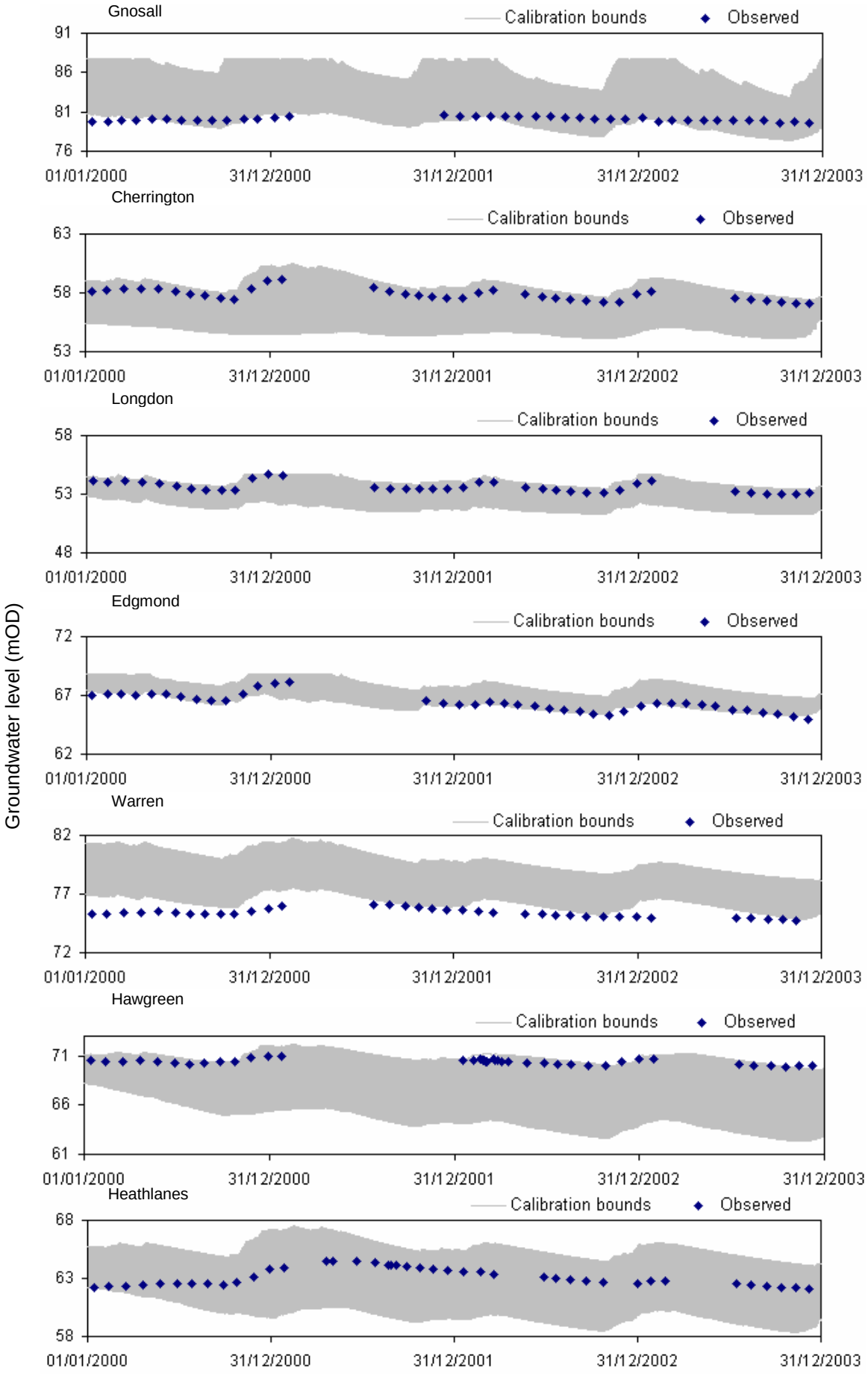


Figure 7.2. Uncertainty bounds in calibrated groundwater levels derived from six calibrated models of different modelling protocols

Table 7.2. Summary score statistics for river flows for the six optimal calibrated models

Spatial Complexity	Calibration method	Measure	Walcot	Roden	Meese	Bailey Brook	Eaton on Tern	Tern at Ternhill	Potford Brook	Strine	**Average**
		Method	Cal	Cal	Cal	Cal	Val	Val	Val	Val	
Homogenous	Manual	Baseflow	0.2	0.22	0.2	0.1	0.19	0.18	0.05	0.19	0.17
		Peaks	0.1	0.15	0.16	0.15	0.12	0.16	0.05	0.15	0.13
		Timing	0.25	0.22	0.22	0.2	0.22	0.22	0.22	0.22	0.22
		MDF	0.2	0.12	0.24	0.05	0.2	0.18	0.02	0.12	0.14
		Total	**0.75**	**0.71**	**0.82**	**0.5**	**0.73**	**0.74**	**0.34**	**0.68**	**0.66**
Homogenous	Autocalibration outlet	Baseflow	0.21	0.18	0.2	0.12	0.2	0.18	0.1	0.19	0.17
		Peaks	0.19	0.17	0.22	0.19	0.11	0.15	0.08	0.14	0.16
		Timing	0.25	0.24	0.23	0.19	0.23	0.22	0.22	0.22	0.23
		MDF	0.22	0.17	0.24	0.07	0.2	0.17	0.04	0.12	0.15
		Total	**0.87**	**0.76**	**0.89**	**0.57**	**0.74**	**0.72**	**0.44**	**0.67**	**0.71**
Homogenous	Autocalibration multi-objective	Baseflow	0.2	0.24	0.15	0.1	0.2	0.16	0.08	0.18	0.16
		Peaks	0.1	0.18	0.22	0.1	0.18	0.19	0.06	0.19	0.15
		Timing	0.23	0.22	0.2	0.2	0.21	0.21	0.22	0.2	0.21
		MDF	0.19	0.12	0.2	0.05	0.21	0.17	0.03	0.12	0.14
		Total	**0.72**	**0.76**	**0.77**	**0.45**	**0.8**	**0.73**	**0.39**	**0.69**	**0.66**
Distributed	Manual	Baseflow	0.21	0.23	0.11	0.1	0.21	0.15	0.08	0.18	0.16
		Peaks	0.11	0.16	0.1	0.2	0.13	0.17	0.06	0.19	0.14
		Timing	0.24	0.23	0.2	0.2	0.21	0.2	0.22	0.21	0.21
		MDF	0.24	0.18	0.24	0.09	0.18	0.15	0.05	0.11	0.16
		Total	**0.8**	**0.8**	**0.65**	**0.59**	**0.73**	**0.67**	**0.41**	**0.69**	**0.67**
Distributed	Autocalibration outlet	Baseflow	0.22	0.22	0.16	0.1	0.22	0.16	0.08	0.21	0.17
		Peaks	0.16	0.21	0.22	0.23	0.21	0.18	0.06	0.18	0.18
		Timing	0.23	0.23	0.2	0.17	0.2	0.19	0.22	0.21	0.21
		MDF	0.23	0.16	0.2	0.12	0.18	0.22	0.04	0.18	0.17
		Total	**0.84**	**0.82**	**0.78**	**0.62**	**0.81**	**0.75**	**0.4**	**0.78**	**0.73**
Distributed	Autocalibration multi-objective	Baseflow	0.21	0.23	0.2	0.1	0.22	0.19	0.09	0.18	0.18
		Peaks	0.19	0.21	0.2	0.2	0.2	0.2	0.09	0.18	0.18
		Timing	0.23	0.22	0.22	0.18	0.2	0.19	0.2	0.18	0.20
		MDF	0.22	0.2	0.24	0.19	0.2	0.2	0.04	0.1	0.17
		Total	**0.85**	**0.86**	**0.86**	**0.67**	**0.82**	**0.78**	**0.42**	**0.64**	**0.74**

Table 7.3. Summary score for groundwater levels for the six optimal calibrated models

Spatial Complexity	Calibration method	Measure	Cherrington	Edgmond	Warren	Gnosall	Hawgreen	Heathlanes	Longdon	Cotton	Radmoor	Rowton	Heathcote	Woodlands	Twinney	Coley	**Average**
		Method	Cal	Cal	Cal	Cal	Cal	Cal	Cal	Val	Val	Val	Val	Val	Val	Val	
Homogenous	Manual	General level	0.46	0.34	0.28	0.45	0.40	0.25	0.48	0.25	0.28	0.30	0.28	0.47	0.24	0.18	0.33
		Overall shape	0.40	0.38	0.20	0.05	0.21	0.35	0.38	0.10	0.25	0.12	0.19	0.47	0.26	0.12	0.25
		Total	**0.86**	**0.72**	**0.48**	**0.50**	**0.61**	**0.60**	**0.86**	**0.35**	**0.53**	**0.42**	**0.47**	**0.94**	**0.50**	**0.30**	**0.58**
Homogenous	Automatic calibration at outlet	General level	0.35	0.25	0.21	0.22	0.38	0.18	0.40	0.21	0.23	0.30	0.24	0.47	0.20	0.19	0.27
		Overall shape	0.40	0.40	0.23	0.13	0.34	0.28	0.40	0.10	0.21	0.10	0.18	0.48	0.25	0.12	0.26
		Total	**0.75**	**0.65**	**0.44**	**0.35**	**0.72**	**0.46**	**0.80**	**0.31**	**0.44**	**0.40**	**0.42**	**0.95**	**0.45**	**0.31**	**0.53**
Homogenous	Autocalibration multi-objective	General level	0.45	0.40	0.32	0.35	0.35	0.30	0.45	0.28	0.31	0.32	0.40	0.42	0.28	0.20	0.35
		Overall shape	0.45	0.40	0.20	0.15	0.35	0.40	0.45	0.14	0.28	0.16	0.32	0.47	0.30	0.12	0.30
		Total	**0.90**	**0.80**	**0.52**	**0.50**	**0.70**	**0.70**	**0.90**	**0.42**	**0.59**	**0.48**	**0.72**	**0.89**	**0.58**	**0.32**	**0.64**
Distributed	Manual	General level	0.30	0.36	0.02	0.01	0.30	0.32	0.28	0.28	0.36	0.26	0.36	0.38	0.30	0.20	0.27
		Overall shape	0.20	0.38	0.15	0.01	0.18	0.20	0.30	0.32	0.35	0.10	0.25	0.22	0.20	0.20	0.22
		Total	**0.50**	**0.74**	**0.17**	**0.02**	**0.48**	**0.52**	**0.58**	**0.60**	**0.71**	**0.36**	**0.61**	**0.60**	**0.50**	**0.40**	**0.49**
Distributed	Automatic calibration at outlet	General level	0.20	0.28	0.04	0.03	0.09	0.20	0.36	0.40	0.34	0.16	0.20	0.36	0.20	0.18	0.22
		Overall shape	0.15	0.39	0.25	0.01	0.20	0.16	0.40	0.34	0.34	0.10	0.20	0.18	0.18	0.20	0.22
		Total	**0.35**	**0.67**	**0.29**	**0.04**	**0.29**	**0.36**	**0.76**	**0.74**	**0.68**	**0.26**	**0.40**	**0.54**	**0.38**	**0.38**	**0.44**
Distributed	Autocalibration multi-objective	General level	0.35	0.40	0.30	0.20	0.38	0.35	0.40	0.25	0.25	0.30	0.37	0.30	0.31	0.20	0.31
		Overall shape	0.20	0.39	0.20	0.10	0.20	0.22	0.40	0.20	0.36	0.09	0.29	0.26	0.21	0.20	0.24
		Total	**0.55**	**0.79**	**0.50**	**0.30**	**0.58**	**0.57**	**0.80**	**0.45**	**0.61**	**0.39**	**0.66**	**0.56**	**0.52**	**0.40**	**0.55**

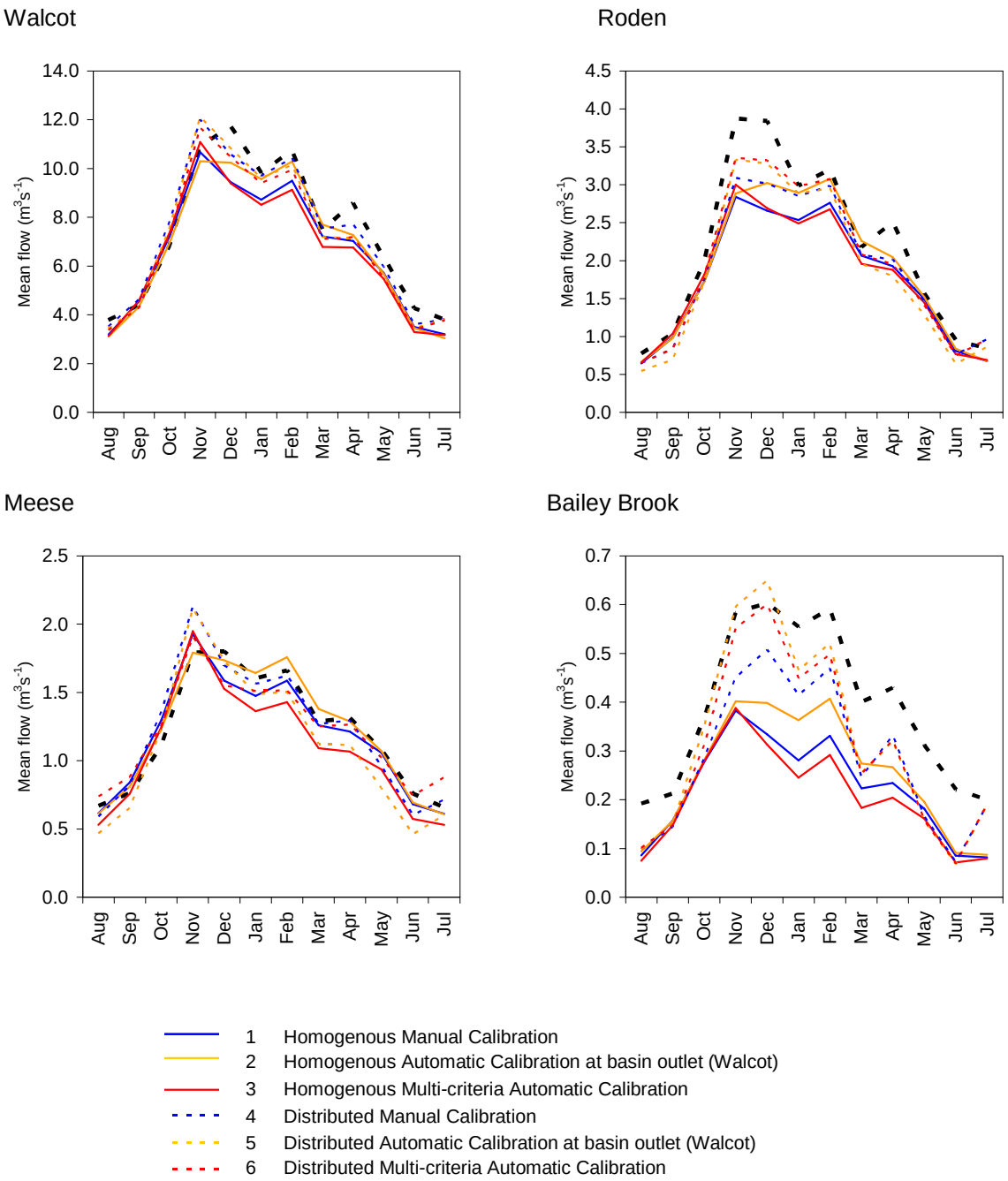


Figure 7.3. Annual river flow regimes (2000-2003). Uncertainty in calibrated river flow in four gauging stations derived from six calibrated models of different modelling protocols

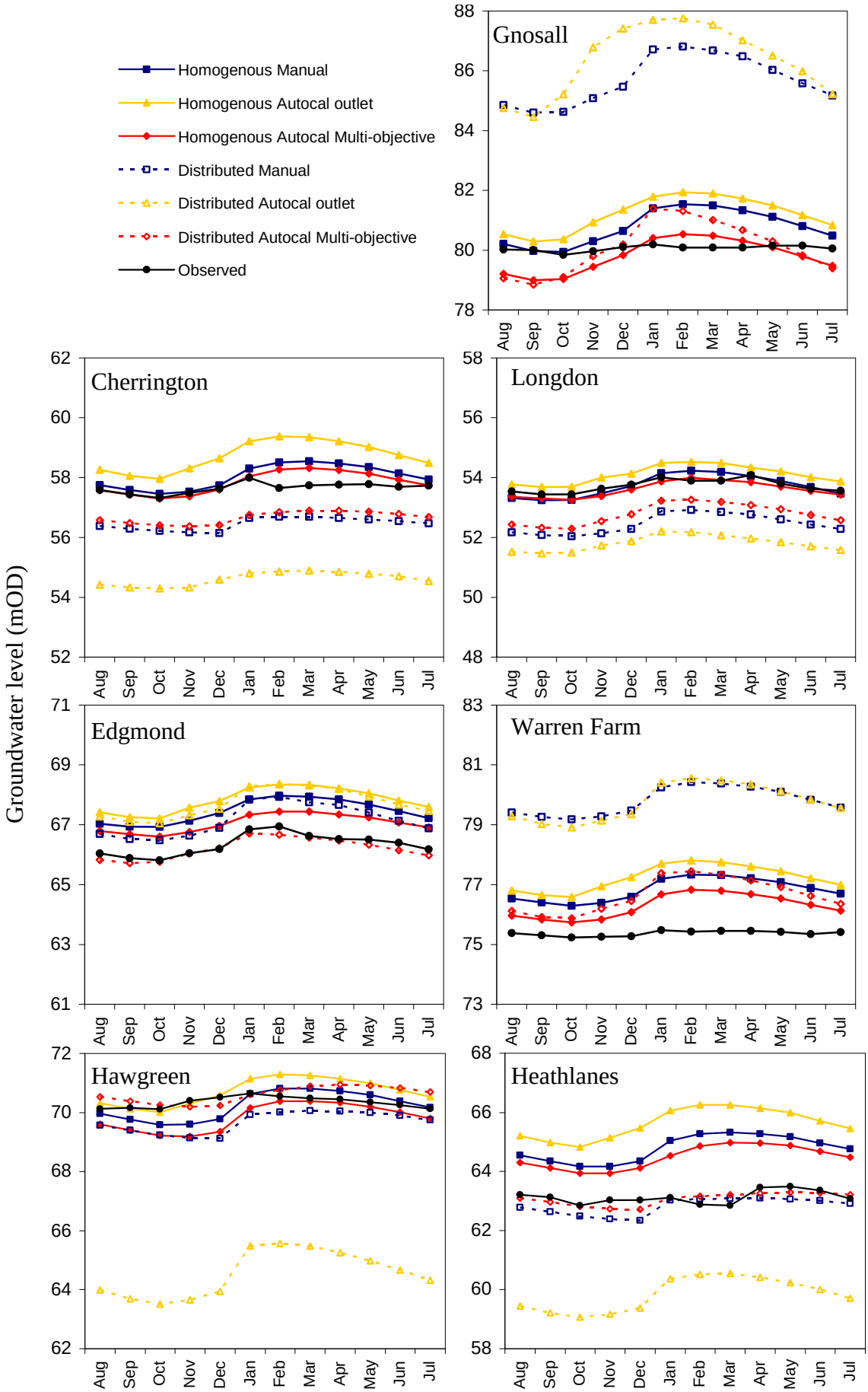


Figure 7.4. Comparative groundwater level regimes (2000-2003) derived from six calibrated models of different modelling protocols

Table 7.4 clearly indicates that statistical performance of river flow at Bailey Brook (calibration) and Potford Brook (validation) are generally classed as very poor (dark grey or orange colour coding) and that performance of river flows at the other gauging stations are generally very good or excellent for range of six models of the different protocols.

The groundwater levels shown in Table 7.5 suggest that at Gnosall (calibration) and Rowton (validation) the performance statistics are not as good as at other sites such as Edgmond (calibration, where all statistics for all modelling protocols are excellent or very good) and Longdon (calibration, where only 3 of all the statistics are not rated as excellent). Performance at Woodlands and Heathcote (both validation sites) is also good, highlighting that the hydrological models do have ability in simulating groundwater levels for sites in which they were not calibrated.

Table 7.7 provides an overall summary of which modelling protocol is statistically the best performing for each of the different statistics, at each of the different river flow and groundwater level boreholes assessed. The calibration and validation sites are identified in the table separately with sites used in the calibration shaded blue, and validation sites shaded green. The table summarises the results of river flow statistics (Table 7.4) and groundwater level statistics (Table 7.5). Values are shown in the table for the sum of the performance for each modelling protocol at each site assessed (river flow or groundwater level). The sum of the performance is calculated according to the coded performance indicator in Table 7.6. One value is shown that comprises of the total score (e.g. 1 to 5 based on the performance code) for all of the statistics calculated for the particular modelling protocol at each site. For example, the score of 21 for Walcot and modelling protocol 1 is derived from summing as score of 4 (very good) for the NSE, 4 for the summary score, correlation and percentage error and a score of 5 (excellent) for the RMSE. As already described, the colour coded performances shown in Table 7.4 and 7.5 relate to the same scoring system as it Table 7.6.

This method of assessing the mean performance for each protocol results in a balanced classification. The alternative of classifying the best modelling protocol according to the highest statistic from each protocol for each site results in other almost equally top performing models being eliminated from the discussion. Table 7.4, for example,

Table 7.4. Compilation of performance statistics for the six optimally calibrated models for river flow

Coded Performance (detailed in Table 7.4) Excellent 5 / Very good 4 / Fair 3 / Poor 2 / Very Poor 1 **Method of calibration**	**Performance criteria**	**Calibration gauges**				**Validation gauges**							**Mean calibration**	**Mean validation**	**Overall mean**
		Walcot	Roden	Meese	Bailey Brook	Roden	Meese	Bailey Brook	Eaton on Tern	Tern at Ternhill	Potford Brook	Strine			
Homogenous Manual calibration	NSE	0.74	0.72	0.75	0.47				0.79	0.70	-0.32	0.60	0.67	0.44	0.56
	Summary score	0.75	0.71	0.82	0.5				0.73	0.74	0.34	0.68	0.70	0.62	0.66
	Correlation	0.81	0.91	0.88	0.85				0.90	0.88	0.70	0.79	0.86	0.82	0.84
	RMSE	2.87	1.27	0.41	0.24				0.74	0.37	0.24	0.37	1.20	0.43	0.81
	% Error	-9.5	-17.8	-2.5	-43.6				7.7	17.4	93.75	9.72	-18.35	32.14	6.90
Distributed Manual calibration	NSE	0.71	0.81	0.5	0.13				0.58	-0.16	-0.18	0.58	0.54	0.21	0.37
	Summary score	0.8	0.8	0.65	0.59				0.73	0.67	0.41	0.69	0.71	0.63	0.67
	Correlation	0.87	0.91	0.85	0.8				0.89	0.85	0.64	0.79	0.86	0.79	0.83
	RMSE	3.07	1.05	0.58	0.31				1.05	0.74	0.23	0.38	1.25	0.60	0.93
	% Error	-1.2	-12.6	0.8	-28.2				21.54	35.24	81.25	13.89	-10.30	37.98	13.84
Homogenous automatic calibration at basin outlet	NSE	0.79				0.72	0.82	0.61	0.78	0.69	-0.26	0.61	0.79	0.57	0.60
	Summary score	0.87				0.76	0.89	0.57	0.74	0.72	0.44	0.67	0.87	0.68	0.71
	Correlation	0.9				0.9	0.9	0.9	0.90	0.89	0.62	0.80	0.90	0.84	0.85
	RMSE	2.57				1.28	0.35	0.2	0.76	0.38	0.23	0.37	2.57	0.51	0.77
	% Error	-7.3				-12.1	0.8	-35.9	7.70	19.05	87.5	9.72	-7.30	10.97	8.68
Distributed automatic calibration at basin outlet	NSE	0.72				0.83	0.53	0.13	0.67	0.23	-0.20	0.60	0.72	0.40	0.44
	Summary score	0.84				0.82	0.78	0.62	0.81	0.75	0.40	0.78	0.84	0.71	0.73
	Correlation	0.88				0.92	0.86	0.83	0.90	0.87	0.64	0.78	0.88	0.83	0.84
	RMSE	2.99				0.99	0.56	0.31	0.92	0.67	0.23	0.37	2.99	0.58	0.88
	% Error	-3.5				-14.5	-8.3	-17.9	9.23	19.05	87.5	5.55	-3.50	11.52	9.64
Homogenous multi-objective automatic calibration	NSE	0.69	0.75	0.63	0.35				0.75	0.60	0.11	0.59	0.61	0.51	0.56
	Summary score	0.72	0.76	0.765	0.445				0.80	0.73	0.39	0.69	0.67	0.65	0.66
	Correlation	0.85	0.91	0.87	0.83				0.89	0.86	0.75	0.79	0.87	0.82	0.84
	RMSE	3.15	1.2	0.5	0.26				0.81	0.44	0.22	0.38	1.28	0.46	0.87
	% Error	-11	-17.8	-10.7	-48.7				10.26	20.95	93.75	9.72	-22.05	33.67	5.81
Distributed multi-objective automatic calibration	NSE	0.71	0.83	0.54	0.13				0.65	0.19	-0.13	0.58	0.55	0.32	0.44
	Summary score	0.85	0.86	0.86	0.67				0.82	0.78	0.42	0.64	0.81	0.67	0.74
	Correlation	0.87	0.92	0.81	0.82				0.89	0.83	0.69	0.80	0.86	0.80	0.83
	RMSE	3.06	1	0.56	0.31				0.96	0.62	0.22	0.38	1.23	0.55	0.89
	% Error	-5.7	-9.8	0.0	-23.1				12.82	20.95	87.5	15.28	-9.65	34.14	12.24

Table 7.5. Compilation of performance statistics for the six optimally calibrated models for groundwater levels

		Calibration boreholes							Validation boreholes															
Coded Performance (detailed in Table 7.4) Excel-lent 5 / Very good 4 / Fair 3 / Poor 2 / Very Poor 1	Performance criteria	Cherrington	Edgmond	Warren	Gnosall	Hawgreen	Heathlanes	Longdon	Cherrington	Edgmond	Warren	Gnosall	Hawgreen	Heathlanes	Longdon	Cotton Farm	Radmoor	Rowton	Heathcote	Woodlands	Twinney	Coley	**Mean calibration**	**Mean validation**	**Overall mean**
Method of calibration																									
Homogenous Manual calibration	Summary score	0.86	0.72	0.48	0.5	0.61	0.6	0.86								0.35	0.53	0.42	0.47	0.94	0.50	0.30	0.66	0.50	0.58
	Correlation	0.81	0.9	0.7	0.4	0.79	0.71	0.89								0.85	0.65	0.10	0.91	0.86	0.41	0.76	0.74	0.65	0.70
	RMSE	0.46	0.91	1.22	0.7	0.83	1.62	0.22								1.77	1.47	0.75	0.86	0.52	1.89	2.13	0.85	1.34	1.10
	% Error	0.7	1.6	2	0.9	-0.2	2.6	0.1								2.80	2.20	0.10	1.01	0.50	3.00	3.00	1.10	1.80	1.45
Distributed Manual calibration	Summary score	0.5	0.74	0.17	0.02	0.48	0.52	0.58								0.60	0.71	0.36	0.61	0.60	0.50	0.40	0.43	0.54	0.49
	Correlation	0.49	0.94	0.65	0.25	0.56	0.22	0.86								0.91	0.69	0.13	0.87	0.58	0.32	0.71	0.57	0.60	0.58
	RMSE	1.35	0.84	4.38	5.69	0.96	0.91	1.3								1.27	0.92	1.64	1.00	1.36	0.89	2.13	2.20	1.32	1.76
	% Error	-2	1.2	5.9	7	-1	-0.5	-2.3								1.96	1.33	1.62	1.21	1.44	0.47	2.98	1.19	1.57	1.38
Homogenous automatic calibration at basin outlet	Summary score								0.75	0.65	0.44	0.35	0.72	0.46	0.8	0.42	0.59	0.48	0.72	0.89	0.58	0.32		0.58	0.58
	Correlation								0.78	0.86	0.63	0.47	0.88	0.66	0.86	0.77	0.73	0.29	0.88	0.83	0.37	0.75		0.70	0.70
	RMSE								1.01	1.45	1.77	1.2	0.53	2.55	0.4	1.98	2.60	0.83	1.31	0.52	2.62	2.14		1.49	1.49
	% Error								1.9	2.2	2.5	1.4	0.4	4	0.8	3.11	3.40	1.10	1.74	0.75	4.44	3.01		2.20	2.20
Distributed automatic calibration at basin outlet	Summary score								0.35	0.67	0.29	0.04	0.29	0.36	0.76	0.74	0.68	0.26	0.40	0.54	0.38	0.38		0.44	0.44
	Correlation								0.43	0.94	0.72	0.25	0.59	0.2	0.83	0.88	0.80	0.49	0.85	0.50	0.22	0.68		0.60	0.60
	RMSE								3.25	1.46	4.79	6.49	6.35	3.71	1.95	0.42	2.95	3.79	2.38	1.22	2.33	2.11		3.09	3.09
	% Error								-5.2	2	5.8	7.9	-8.3	-5.2	-3.5	0.02	4.28	5.91	2.99	0.93	3.36	2.98		1.00	1.00
Homogenous multi-objective automatic calibration	Summary score	0.9	0.8	0.52	0.5	0.7	0.7	0.9								0.42	0.59	0.48	0.72	0.89	0.58	0.32	0.72	0.57	0.64
	Correlation	0.8	0.77	0.63	0.35	0.79	0.82	0.79								0.83	0.63	0.07	0.90	0.73	0.39	0.75	0.71	0.61	0.66
	RMSE	0.44	0.81	0.9	0.7	0.89	1.41	0.29								1.69	1.35	0.59	0.42	0.88	1.37	2.10	0.78	1.20	0.99
	% Error	0.3	1	1.2	-0.3	-0.8	2.2	-0.1								2.84	2.18	0.16	0.49	1.10	2.84	2.97	0.50	1.80	1.15
Distributed multi-objective automatic calibration	Summary score	0.55	0.79	0.5	0.3	0.58	0.57	0.8								0.45	0.61	0.39	0.66	0.56	0.52	0.40	0.58	0.51	0.55
	Correlation	0.47	0.86	0.51	0.1	0.52	0.12	0.91								0.89	0.89	0.33	0.84	0.59	0.19	0.74	0.50	0.64	0.57
	RMSE	0.63	0.88	0.64	0.3	0.69	0.46	0.86								1.61	2.00	1.02	1.37	2.67	0.81	2.12	0.64	1.66	1.15
	% Error	-1.7	-0.2	1.7	0	0.3	-0.1	-1.7								2.75	2.81	0.21	1.71	3.85	0.75	3.00	-0.24	2.15	0.96

demonstrates that the correlation statistic is very similar for the six protocols for the Roden tributary with the statistic ranging between 0.90 to 0.92 for all sites. On the contrary the method that has been selected classifies the statistical performance (in this case all would be scored 5 - excellent). In this way a total score is derived for each site and protocol and equally values each of the different statistics.

Table 7.7 is further discussed at the end of this chapter (Section 7.4), as a means of numerically assessing whether to accept or reject the decision if any of the modelling protocols can be considered better than the rest.

Table 7.6. Model performance classification for each performance criteria

Performance Criteria	**Excellent**	**Very good**	**Fair**	**Poor**	**Very poor**
Coded Performance	5	4	3	2	1
Nash-Sutcliffe, NSE, a	> 0.85	0.65 – 0.85	0.50 – 0.65	0.20-0.50	<0.20
Summary Score, c	> 0.85	0.65 – 0.85	0.50 – 0.65	0.20-0.50	<0.20
Correlation, R, c	> 0.85	0.65 – 0.85	0.50 – 0.65	0.20-0.50	<0.20
RMSE (m^3s^{-1} or m), a	<4	4-6	6-8	8-10	>10
% Error, b	<5	5-10	10-15	15-20	>20

a – Henriksen et al., 2008, b – Andersen et al., 2001, c – Using same criteria as for Nash-Sutcliffe, a

Table 7.7. Mean performance of the different protocols according to each performance measure for river flow gauging station and groundwater level boreholes for calibration and validation sites

1 Homogenous Manual Calibration
2 Distributed Manual Calibration
3 Homogenous Automatic Calibration at basin outlet (Walcot)
4 Distributed Automatic Calibration at basin outlet (Walcot)
5 Homogenous Multi-criteria Automatic Calibration
6 Distributed Multi-criteria Automatic Calibration

	River flow											Groundwater levels (GWL)																Summary		
Model protocol	Walcot	Roden	Meese	Bailey Brook	Eaton on Tern	Tern at Ternhill	Potford Brook	Strine	**Calibration mean (river flow)**	**Validation mean (river flow)**	**River flow mean (calibration &validation)**	Cherrington	Edgmond	Warren	Gnosall	Hawgreen	Heathlanes	Longdon	Cotton Farm	Radmoor	Rowton	Heathcote	Woodlands	Twinney	Coley	**Calibration mean (GWL)**	**Validation mean (GWL)**	**GWL mean (calibration & validation)**	**Mean calibration (river flow & GWL)**	**Mean validation (river flow & GWL)**	**Overall mean**
1	21	20	24	16	22	20	13	18	20.3	18.3	19.3	19	19	16	15	17	17	20	17	17	13	17	20	15	16	17.6	16.4	17.0	18.5	17.1	17.8
2	24	21	23	16	18	16	12	19	21.0	16.3	18.6	15	19	13	15	15	15	18	18	18	13	18	16	15	16	15.7	16.3	16.0	17.6	16.3	17.0
3	24	21	24	17	18	20	12	20	24.0	18.9	19.5	18	19	15	14	19	16	19	16	17	14	19	19	15	16	na	16.9	16.9	24.0	17.5	17.8
4	24	21	21	19	22	18	12	20	24.0	19.0	19.6	13	19	14	14	12	13	18	16	18	13	17	16	14	16	na	15.2	15.2	24.0	16.5	16.8
5	21	20	20	14	21	18	13	20	18.8	18.0	18.4	19	18	16	5	18	18	19	19	16	13	19	19	15	16	16.1	16.7	16.4	17.1	17.2	17.1
6	24	24	23	15	21	15	13	17	21.5	16.5	19.0	15	19	16	13	16	14	19	17	18	14	18	16	18	16	16.0	16.7	16.4	18.0	16.6	17.3
mean site score	23.0	21.2	22.5	16.2	20.3	17.8	12.5	19.0	21.6	17.8	19.1	16.5	18.8	15.0	12.7	16.2	15.5	18.8	17.2	17.3	13.3	18.0	17.7	15.3	16.0	16.4	16.4	16.3	19.9	16.9	17.3

The scores are based on the performance shown in table 7.6 where each performance statistic was rated as excellent if a score of 5 was given, if it was poor then a score of 1 was given. The totals shown here are made up of the sum of the scores for all the statistics for each site (e.g. for river flow sites: NSE, R, Summary Score, RMSE and % error (max score of 25) For GW boreholes: R Summary score, RMSE and % error (Max score of 20))

Cells shaded in blue refer to calibration sites. Cells shaded in green refer to validation sites.

7.3. Assessment of thesis research questions

7.3.1. Model spatial complexity

As introduced in the discussion of different types of hydrological model in Section 2.2, models of varying spatial representations; lumped catchment hydrological models, semi-distributed models and distributed models have all been developed that can produce similar ability in simulating river flows at the basin outlet (e.g. Boughton (2006), Jiang et al., (2007), Kite (1995) and Sahoo et al., (2006)). Most of the published research that addresses the issues of model spatial complexity compare models that have been constructed using different model codes. For example Abu El Nasr et al., (2005) compared MIKE SHE and SWAT model codes in the 465km^2 Jeker Basin, Belgium, whilst Yang et al., (2000) modelled the 703 km^2 Seki Basin in Japan using MIKE SHE, TOPMODEL and the GB model. Although the method of assessing spatial complexity using different model codes is appropriate for assessing reproducibility of results with different codes, it introduces further uncertainty in directly comparing the issue of how the representation of different spatial complexity affects model output through differences in model structure uncertainties.

The issue of representing a catchment with different spatial representations in hydrological models, and comparing the differences in outputs from these different models was reviewed in Chapter 2. It is of importance for a variety of reasons including implications for the scale at which data collection and monitoring is undertaken, as well as implications for the time and effort requirements for human resources needed for more complex spatially distributed modelling.

In order to compare intra-model code spatial representation in model results, the MIKE SHE code was used to represent processes and variables at different spatial resolutions. A homogenous model (with mostly uniform parameter values across the catchment), and a model where the representation of variables and parameters were more comprehensively spatially distributed were constructed, tested and calibrated in the chapters 4 - 6. In all stages of the modelling process, the models of the two different spatial complexities were subject to the same construction and testing methodology to ensure the results could be comparable.

Aside from small differences, the simulated general annual flow regimes for the six models shown in Figure 7.3 are very similar – especially at the basin outlet, Walcot, for

both the homogenous (solid line) and distributed (dashed line) models. The percentage error between observed and simulated models range from -1.2% to -11% (Table 7.4) which fall in the category of excellent and very good (Table 7.5). Both complexities (and all calibrations) of models simulate flow at Walcot, in the Roden and the Meese with flow regimes of peaks (November) and secondary peaks (February) and low flows (June-August) in the same months. The similarity of simulated river flows at Walcot between the six models is discussed further in Section 7.3.5 which addresses issues of model assessment at the outlet as well as internal performance measures, especially with regard to model scaling. Based on the results at the basin outlet, however, neither the homogenous or distributed protocols show any distinctive patterns to support one as better than the other.

The simulated annual flow regimes at Bailey Brook, the smaller upstream tributary, show larger differences between model simulations especially in winter months between November-February. The homogenous models are shown to simulate a substantially reduced winter river flow when compared to the distributed models. The differences between the models simulations in Bailey Brook may be a result of the homogenous model using uniform values for the majority of parameters in the model. It is likely that the calibrated parameter sets may not be best representative for this specifically mudstone dominated sub-basin. This is highlighted, for example, by the horizontal hydraulic conductivity value used being best representative of sandstone, and as such it is likely the observed peaks are not simulated as overland flow is reduced. Conversely, the distributed model includes more sub-basin specific data, such as representative geology or input precipitation deriving from the Sandford gauge (Figure 6.3) rather than the catchment mean which is used in the homogenous model.

Figure 7.4 displays the observed and simulated annual groundwater level regimes at seven boreholes for the six calibrated models. For each borehole the y axis has been fixed to a scale of 10m to enable comparisons between the sites and model outputs. As shown, the shape of the regimes are similar at all boreholes for both homogenous and distributed models, with winter recharge evident between November and February, and summer draw down occurring between May to September. Despite the similar shapes of the annual groundwater level regimes, the mean groundwater level (reflected by the widely varying RMSE (m) and % Error in Table 7.4) is the characteristic that is most variable between the homogenous and distributed models. Although the calibrated

range (the uncertainty between models) appears large, the magnitude in the differences in mean groundwater levels are in the same order as other research simulating groundwater levels at the catchment scale (eg. Madsen, 2003; Henriksen et al., 2008).

When compared to the observed groundwater levels, the various calibrations of the homogenous models appear to result in better simulations of the general groundwater levels, with RMSE (m) (Table 7.4) lower than for the distributed models. This is also reflected by the larger number of solid lines (which relate to the homogenous model) in Figure 7.7, where colour codes are shown in the table to represent the model with the best performance according to each performance criteria.

It is expected that the results for simulated groundwater levels with the distributed model (with the exception of the multi-objective automatic calibration, discussed in Section 7.3.2) are of a poorer standard for a number of reasons. The sub-surface of the Tern Catchment in reality is a very complex system with widely varying drift deposits of different natures and thickness, as well as substantial differences in solid geology stratigraphy. Although Chapter 4 highlighted that the distributed model focussed widely on spatially varying soil and solid geology, the trade off between grid size and computational processing time, data availability and importance of reducing parameterisation may be a main cause of the distributed model outputs often being poorer than the homogenous model, that sought an average parameter set for the catchment. Although Figures 3.5 and 3.6 highlighted that the data for the spatial variation in geology is available at a higher resolution than the 1km grid size used within the models, data for parameter values that are associated with the different geology types were not available and so were subject to calibration, with the values ascertained from the literature (Hobbs et al., 2002, Johnson, 1967, Allen et al., 1997 in LOCAR, 2000). The issue in Section 2.3.3 of up-scaling parameter values to represent larger areas from which they were measured remains a fundamental problem for distributed hydrological modelling, to an extent that it limits the credibility of the distributed models (Moreda et al., 2006).

The results of this research support the suggestion that the objective of the research should define the level of spatial complexity used in a hydrological model (Refsgaard, 1997). Yang et al., (2000) also suggest that ‘the complexity of the hydrological models

and their flexibility for representing spatial heterogeneity determine them as suitable for different purposes'. Additionally, if the input data is not of high enough spatial resolution (or the parameters without measured data not represented very well by the values used) it may be better to use a simpler/homogenous model. It is important to recognise that a large amount of human bias and more uncertainty can be added by the hydrological modeller when attempting to average parameter values within a grid cell at a size larger than it is representative.

If a spatially distributed model is considered necessary in order to fulfil the research objectives - such as in research that specifically tracks pollution or nutrient movement within the catchment, (e.g. Brun et al., 2002a and b; Christiansen et al., 2004) then consideration also needs to be given to the time needed for setting up a distributed model. In this research, the construction, calibration and processing of the distributed model took considerably more time than the homogenous model. The results indicate that if the purpose of a piece of research is to assess river flow at different locations within a catchment, then a model more homogenous in nature can result in very similar model performance.

It is important to note that this research has been undertaken at a catchment scale and compares overall performance within the whole catchment, and therefore these results may only be specific to the Tern Basin (with its complex geology) and/or to catchments of a similar size using a similar grid size. Further research could test whether in a smaller scale study, with a higher resolution of data the results may be different. As Yang et al., (2000) conclude, in the 703 km^2 Seki catchment, Japan, MIKE SHE may be most useful for comprehensive hydrological simulations in small catchments. For example, smaller wetland modelling studies such as Thompson et al., (2004) and Thompson et al., (2009), require large amounts of data, but at these smaller scales it is in theory possible to collect. In such cases, the concept of a reliable and operational distributed model is more likely.

The uncertainties apparent from the parameterisation and scaling of the distributed model at the catchment scale in this research therefore highlights the question 'at what scale is a model able to be called truly spatially distributed' and if, as Beven (1989) notes, is it really possible? In theory it is possible to construct a spatially distributed

model, and as Silberstein (2006) suggests, the mathematical models have likely been developed enough for this cause. However, at a mid-size basin-scale, or catchment scale, its application is unrealistic as there is a lag in the amount of data that need to be collected, the time this takes (to collect and process, as well as simulate in a model with small enough grid sizes or spatial resolution) is too great.

7.3.2. Model calibration

Another of the main research themes has been to assess calibration uncertainty on simulated river flow and groundwater levels at a selection of sites within the catchment. The need for calibration in hydrological models (Section 2.6) was shown to be almost unavoidable (Refsgaard, 1997), and different methods of model calibration were reviewed. Manual and automatic calibration methods for hydrological models were discussed, and aside from suggestions that distributed models need more rigorous calibration methods due to the increase in parameterisation (Moreda et al., 2006), little research directly compares simulation results derived from different methods of calibration – especially whilst simultaneously comparing models of different spatial complexities and assessing results at a number of internal gauging stations and boreholes within the catchment.

The issue of a number of models being calibrated using different calibration methods is different from that of equifinality, where many models can be considered calibrated (as also addressed in the thesis) as these models typically derive from the same method of calibration but use different parameter sets (for example using GLUE or an automatic calibration methodology (Beven and Binley, 1992)).

Both the homogenous (Chapter 5) and distributed (Chapter 6) hydrological models of the Tern Catchment were first manually calibrated. This stage is shown in the highlighted boxes at the top of Figure 7.5 (taken from Figure 1.2 that provided a conceptual strategy for the research).

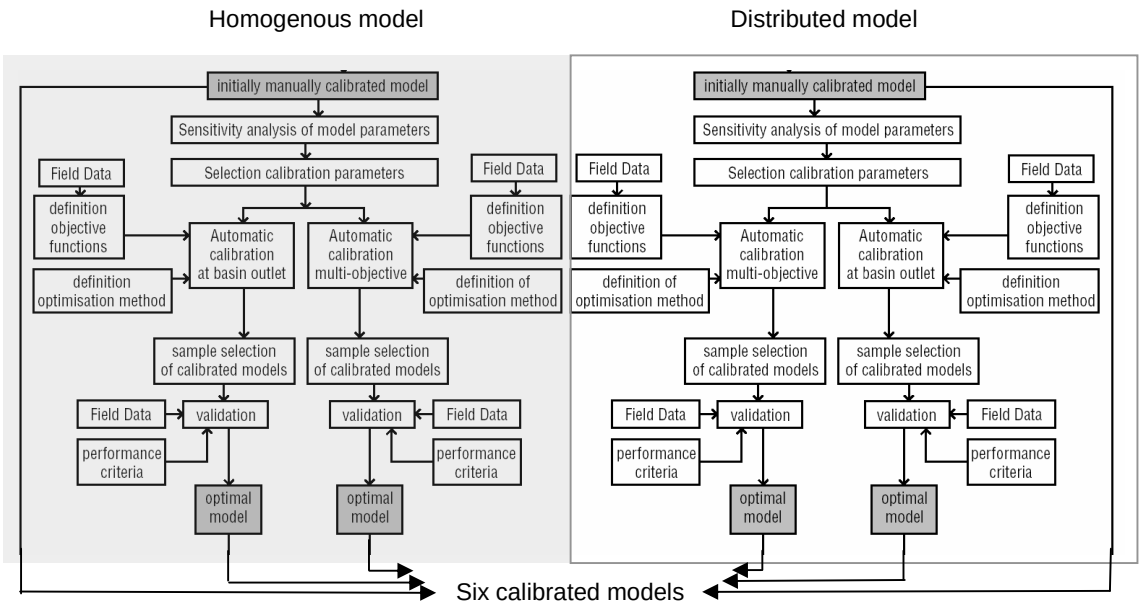


Figure 7.5. Method of calibrating and selecting six optimal calibrated models from different modelling protocols

The process of manual calibration was undertaken using the same method for both models – with key parameters adjusted manually with the simulation re-run and results extracted. Parameter values were modified so as to fulfil two criteria; being values within limits specified so as to be representative and realistic for the catchment, and also for the simulation to result in the best fit possible with observed flow and groundwater levels. Simulation statistics (Nash and Sutcliffe NSE, Correlation R and RMSE) were also used as a quantified measure of model performance.

Both of the hydrological models were then used in two further methods of calibration; an automatic calibration that using performance criteria (RMSE) optimised river flow only at the basin outlet, and the other using a multi-criteria approach similar to Madsen (2003), that optimised model performance (RMSE) using a range of observed river flow and groundwater levels at internal locations within the catchment. e The methodology of this stage is summarised for both homogenous and distributed models in Figure 7.5. The issue of equifinality (discussed in the next section) resulted in the presentation of results for a suite of ten calibrated models for each automatic calibration method (in chapters 5 and 6). However, the automatically derived optimum models for each automatic calibration method (highlighted in the boxes at the bottom of Figure 7.5) are used as the six models by which assessments of calibration methods (and spatial complexity issues) are discussed.

Tables 7.4 to 7.7 that highlight the performance statistics and performance evaluation of the calibrations of the different models show overall fair calibration and validation statistics. For example, Table 7.4 shows the mean river flow NSE calculated for the river flow gauging stations for the six models range between 0.54 to 0.79 for the calibration sites and 0.21 – 0.57 for the validation sites (lower due to the poor performance in Potford Brook). The corresponding range for the groundwater level boreholes shown with the mean correlation coefficient is 0.50 – 0.74 for the calibration period and 0.60 – 0.70 for the validation period.. This performance can be categorised as fair to very good (Table 7.3). Evaluation of the results at individual gauging stations and boreholes are shown (Tables 7.4 and 7.5) to range more widely, with some sites classed as excellent and others as poor for the range of statistics. This result is similar to other research for example Refsgaard, (1997) and Madsen (2003) which also highlight a similar finding that in no study are all internal sites where model output is assessed shown to have equal performance.

The differences in colours shown in Figures 7.3 and 7.4 reflect the different calibration methods used for the models. It is shown in Figure 7.3 for river flow gauging stations, that the automatically calibrated models at the basin outlet (yellow lines) appear to show a better fit with the observed data, especially during winter months, when compared to the other calibration methods. This is expected to an extent as this calibration method optimises solely on river flow performance. Statistically this is also supported as highlighted in Table 7.7. The mean calibration score (24 for both protocols 3 and 4) and mean validation score (18.9 and 19 for protocols 3 and 4) are the highest when compared to the other calibration methods in the other protocols.

In addition to the automatically calibrated models, it is notable that Table 7.4 shows that the manually calibrated models (blue lines/protocols 1 and 2) are rarely the best performing models according to each of the different statistics for river flow. In reporting thoroughly on the process of calibration (not just the quantitative results), there was a feeling of uncertainty when undertaking the manual calibrations, especially for the distributed model. For example, this method of calibration depends on conceptual understanding of the catchment, the processes operating within it, and the understanding of the mathematics of the model. In the distributed model which included many more parameters than the homogenous model, manual perturbation of individual

values with the objective of improving performance was more difficult when relationships between parameters were not as well known, and for this reason it is difficult to justify full confidence in the calibration. This is also supported by the lowest mean validation score for river flow in Table 7.7 being a score of 16.3 for protocol 2 (the manual calibration of the distributed model).

In consideration of the differences in calibrated groundwater levels shown in Figure 7.4, none of the calibration methods appear to result in a substantially better performance than the others. However, it is notable that some of the simulated groundwater levels of the model automatically calibrated using river flow at the basin outlet are exceptionally poor in comparison to the other calibrations. This is confirmed in Table 7.7 with protocol 4 (distributed model automatically calibrated for river flow at the basin outlet) having the lowest mean score for groundwater levels of all the protocols with a value of 15.2. Also, at Heathlanes, Hawgreen, Warren Farm and Gnosall the RMSE(m) values (Table 7.4) are particularly high compared to those derived as a result of other calibration methods. Figure 7.4 shows that levels at these locations often differ by several metres compared to those from the other calibrated models. This result is important in highlighting that there is a need for a robust calibration methodology, and if internal performance measures are not included during the calibration process (i.e. the model is only calibrated at the outlet) then internal processes may be simulated very poorly despite the being calibrated well.

It is therefore important to look at the validation statistics for the more comprehensive multi-criteria calibration methods to assess whether they indicate a more robust method. This is confirmed in Table 7.7 when noting that for both protocols 5 and 6 that are automatically calibrated models with the multi-criteria approach show an increase in the mean validation score when compared to the calibration score. The values increase from 16.1 and 16.0 to 16.7 for protocols 5 and 6, respectively. Based on the mean performance score, the validation results being higher than those that were calibrated indicates the models perform successfully outside the sites used in calibration.

These results suggest that the method of calibration is important in hydrological modelling, mostly for model reliability and confidence issues in the way the model simulates different processes. The work that has been undertaken in this thesis

demonstrates that although different calibrations can result in similar performance statistics (at the basin outlet especially), the calibration methods have comprised of a range of different complexities that infer ranging degrees of confidence in the ability to simulate catchment processes. It has not been possible to firmly select a best calibration method, it is only possible to note that according to the validation statistics in Tables 7.4, 7.5 and 7.7 that the multi-criteria automatic calibration method are the most robust of those tested.

7.3.3. Parameter uncertainty

An important sub-section of the calibration and validation issues detailed in Section 2.6 included the discussion of parameter uncertainty and equifinality. As previously described, issues of equifinality within the parameter space have been addressed in the preceding chapters 5 and 6, by the inclusion and selection of sets of ten calibrated models for each of the automatic calibration methods. As Beven and Binley (1992) showed using the GLUE methodology and TOPMODEL, it has also been shown using MIKE SHE and an independent method in this thesis, that different parameter sets can produce model simulations with performance statistics that are very similar.

In this research where uncertainty in the modelling process is being addressed (especially in the model calibration phase), the issue of parameter equifinality was considered as an issue too important to eliminate from the research and so has been explored. However, it is not the main issue being addressed in the thesis, and, as already described only the six selected models of different spatial complexities and model calibration methods are discussed in this chapter.

7.3.4. Quantitative/statistic model performance

A further research theme of the thesis has been to assess how performance statistics that are commonly used in hydrological modelling suggest different abilities of the models. Throughout chapters 5 and 6, this objective was addressed by first describing the ability of the models from the results of river flow and groundwater levels at different locations within the catchment. This expert elicitation has then been quantified with the calculation of a further measure; the summary score, that qualitatively (with the exception of the mean daily flow rated the ability of the model at simulating observed base flows, peak flows, the timing and difference in lags as well as the differences in

mean daily flows. The commonly used correlation (R), Nash-Sutcliffe (NSE), Root Mean Squared Error (RMSE) statistics were also calculated for each model at various sites within the catchment.

Given the review of performance statistics in Section 2.6.2, it was considered important to assess how different statistics reflect different model performance as there was an identified gap in the literature where this has not previously been undertaken in detail. It was acknowledged that a large number of statistics exist, and that different statistics may be of use for assessing different parts of the hydrograph or model output in comparison to observed data, but in practice, there is little research in physically-based distributed modelling that compare different performance measures for models constructed using the same modelling code and the same catchment.

In this research, performance statistics have been calculated at 22 locations within the catchment – eight river flow gauging stations and fourteen groundwater level boreholes as shown in Tables 7.4 and 7.5. As performance statistics have also been calculated for each of the six modelling protocols a large quantity of performance data can be compared to see if statistical consistency is shown. For example, it is possible to assess whether any of the output measures (e.g. river flow at a particular gauging station) are consistently calculated by performance measures to consistently indicate a best or worst performance.

Table 7.4 provides a summary of all the statistics calculated for six optimal simulations from each protocol for river flows. As consistently shown for the mean river flows for all six protocols, the R, NSE and RMSE show the best performance for each statistic for the same model – the homogenous model automatically calibrated at the basin outlet, where values of 0.90, 0.79 and 2.57 are shown (Table 7.4).

For mean groundwater levels (Table 7.5 and Table 7.7), the performance of the different statistics are in contrast to those for river flow. The R, RMSE and summary score each suggest that different modelling protocols show the best performance. As a result, it is difficult to suggest quantitatively that any one of the six models is better than any other at simulating groundwater levels. A correlation, R, of 0.74 the highest value for the homogenous manual calibration, 0.64 the best RMSE for the distributed model automatically calibrated using the multi-objective method, and the summary score was

highest for the homogenous model automatically calibrated using the multi-objective method with a score of 0.72.

In discussing the different performance statistics it is also notable that in general the mean score for calibration (for river flows and groundwater levels together) is higher than for the mean validation score, with values of 19.9 and 16.9 respectively. This indicates that in general, for the protocols assessed, the model does not perform as well when tested against observed data that was not used in the calibration process. Although this is not ideal, it is often shown in hydrological modelling (Refsgaard, 1997) and in this case the reduction in the score is not so large as to merit a large problem.

The reasons for the discrepancies between output statistics indicating different modelling protocols that perform the best may be due to the degree in which the particular statistic has been included within the calibration of the model and the calibration method used (detailed in the subsequent paragraph). It is suggested that this in turn should be determined by the objective and purpose of the modelling study. The performance statistics that are used may play a larger part in how well a model is calibrated rather than the statistic only being used as a tool to quantitatively assess performance.

To explain this further, the different calibration methods and the way the different statistics were used within each calibration method are discussed. During the manual calibrations for both the homogenous and distributed models, particular attention was given primarily to how well simulated river flow and groundwater levels reproduced the observed data (from which the summary score was calculated). Additionally, the different performance statistics, notably the NSE for flows and R for groundwater levels were reviewed, in order to maximise their values whilst at the same time trying to ensure that the model parameters used in the calibration were realistic and representative of the catchment. As the manual calibration process included many output measures with different statistics for each, it was hard to focus on the values of the statistics and on all the measures. Consequently, this is perhaps why the highest groundwater level R, is found for the homogenous manual calibration, as relatively little attention was given to the RMSE in this stage.

For the automatic calibrations of the homogenous and distributed models that were optimised using the output measure of RMSE of river flow at the basin outlet, it was expected (and shown in both Chapters 5 and 6) that the resulting RMSE values of river flow should be best for this calibration method. The optimisation of the calibration was not biased by using other internal measures and so resulted in a model that although apparently good at simulating river flows at the basin outlet, was less able to simulate groundwater levels (shown by the decrease in the performance of R, RMSE and the summary score at the groundwater level boreholes when compared to the manual calibration values).

Lastly, the automatic calibrations that used an objective function that equally weighted the river flow RMSE and groundwater level RMSE at a number of sites again focussed solely on the RMSE statistic optimisation. Although this calibration method is arguably less biased, the RMSE is a statistic that is concerned with the mean error, and so is not specific to assessing parts of the simulation that may be most important (such as if specifically aiming to simulate flood events from peak flows – then the rest of the hydrograph would also be assessed equally with this statistic).

The purpose and aim of any research that uses a hydrological model that requires calibrating should be considered when deciding not only what spatial representation and complexity of model to use, but also what statistics to use in the process of calibration and testing, as now described with examples from the literature. An integrated modelling study that simulates groundwater levels using a conceptual representation of the sub-surface (such as using the linear reservoir method within MIKE SHE (DHI-WE, 2005)) is used as the first example. The main purpose of the research in such a study may be concerned with processes of stream-flow, flooding and runoff response such as Sahoo et al., (2006) where a detailed groundwater model is not considered necessary. In this case, it would be important for the hydrological model to be able to accurately simulate peak flow and response times. In the process of model calibration, it may therefore be decided that manual calibration of the NSE statistic coupled with inspection of the output hydrographs would be adequate. The NSE statistic would then be used within the calibration procedure, with emphasis both on maximising the NSE value whilst balancing the model parameters used in the calibration so as to remain realistic in their physical representation.

If the purpose of a study is to simulate detailed groundwater levels, such as Henriksen et al., (2003) that developed a model to simulate groundwater levels on a 1km × 1km grid scale for a 7330km^2 island in Denmark, a more detailed representation of the sub-surface using the 3D finite difference method (DHI-WE, 2005) would be the appropriate method of process representation. In this case a more detailed method of model calibration may be considered, such as automatic calibration that weights a number of different measures and concentrates more fully on the groundwater levels than surface flows.

The consistent approach to calculating a range of statistics at a range of locations has shown that the values of different statistics used in the assessments can vary considerably even between individual sites. The statistical ability of the model at simulating groundwater levels at Warren Farm range from R of 0.51 to 0.72, which although not as high as for other boreholes, is still a fair performance. The result highlights the weakness of the correlation, R, statistic that if used independently would result in a large error in judgement of model ability. It was demonstrated in chapters 5 and 6 that the simulated levels were poor, especially at simulating the general level. It is expected that this is why the R statistic is defined as acceptable yet the RMSE values compared in Table 7.4 are poor, ranging from 0.64m at best, to 4.79m at worst, as they consider the mean error in simulated levels. These results therefore highlight the need to be cautious about which statistics to use in model assessment, and as already described for what purpose the research is being undertaken.

The need for independent review of the model output by the modeller is therefore clearly required. The quantitative statistics are shown as good tools in assessing model performance, but it is also shown that the quantitative statistics can often be mis-leading or confusing when trying to make an overall assessment with more than one statistic. Qualitative model assessment or expert elicitation (Refsgaard et al., 2006) has been undertaken rigorously in the research. Initially this was achieved by qualitative account of model performance and later improved by inclusion of the summary score as documented throughout Chapters 5 and 6, and shown as a summary in Tables 7.2 and 7.3 and Tables 7.4 and 7.5. The resulting summary score measure has enabled a

methodical process to qualitative model assessment that has been documented to show comparable results throughout the different calibration methods in the previous chapters. The results of this thesis therefore suggest that an independent assessment of performance needs to be undertaken using a separate statistic rather than only assessing statistical performance with performance measures used in the calibration process.

7.3.5. Assessment of internal model performance

As already reviewed, the internal performance of the hydrological models have been consistently assessed at four gauging stations and seven groundwater level boreholes. Where necessary, a further four internal gauging stations and seven groundwater level boreholes were also used as validation sites to test the performance of the model calibrations. As Refsgaard (1997) illustrated, and as has been demonstrated in this work, there is a need for incorporation of multi-criteria and multi-scale assessment of distributed hydrological models. Multi-criteria assessment and review of internal model performance is used in the thesis to address scaling issues as well as to assess whether any trends in the model output response from each of the different calibrated models would be the same.

Throughout the thesis model assessment has consistently been undertaken at a range of locations; from river flows at the basin outlet as well as on major and minor upstream tributaries, as well as for groundwater levels located in different geologies. As discussed by Beven (1989) and Refsgaard (1997), this internal assessment is necessary as distributed models (and in this case homogenous models set up on a distributed gridded system) can have complex internal responses that if not tested and validated can result in many of the simulated processes being poorly represented.

In Table 7.7 it is shown that best performance in the Roden tributary derives from the distributed model with multi-criteria automatic calibration – protocol 6 (score of 24), a notable result as in almost all other sites, the homogenous model statistically performed better. The result is important as it highlights that the parameter set that is best calibrated at the outlet may not always be the best for representing small sub-areas within the catchment. It is expected that the distributed models indicate better results in the Roden tributary (when compared to the Meese, for example) because of the

differences in the solid geologies and difference in soil types of this tributary. The distributed models seek to account for some of this spatial variation whereas the homogenous models do not.

The individual statistics calculated for each of the internal performance measures in Tables 7.4 and 7.5 clearly suggest that some sites have better statistical performance than others. For example, the NSE statistic ranges between 0.69 – 0.79 for river flow at the basin outlet (from the range of the six models), a performance rated as very good according to the criteria in Table 7.6 (Henriksen et al., 2008). Despite this, the performance at the upstream tributary Bailey Brook ranges between an NSE of 0.13 – 0.61 (performance rated between very poor to fair). These results are typical of distributed model performance and do not indicate a new problem, however, results of this research further review the problem by highlighting that despite having used a range of methods of calibration and spatial complexity in data representation, the issue is not overcome.

Additionally, the differences between the sub-basin simulations for the homogenous and distributed models at Bailey Brook (shown in Figure 7.3) may be a result of scaling factors in the model. The grid size of 1km × 1km was chosen because of a necessary trade off between computer processing time and the resolution of available data. This grid cell size is comparable to that used in other studies (e.g. Table 2.3).

In Bailey Brook each 1km × 1km grid cell represents 2.9% of the total sub-basin area. In contrast, the same grid cell represents 0.12% of the whole Tern catchment. In Bailey Brook therefore, each cell averages more of the spatial complexity than equivalent for the whole catchment. Additionally, the river flow simulated at the outlet of Bailey Brook is only comprised of water in the small upstream area (and is therefore sensitive to the parameter values representing that area). However, at the basin scale the simulated river flows result from runoff derived from a larger model area, and so whether the parameter values are specific over a small individual area (such as Bailey Brook) cannot be seen at the basin outlet due to the averaging out of the data.

Compared to the averaging out theory of river flows as shown at the basin outlet, any groundwater level borehole reflects the surrounding parameter values more precisely as

boreholes are point specific and reflect more closely the parameter values in surrounding and the specific grid cell they represent (as movement between adjacent cells is also accountable for the groundwater level).

Therefore, assessing the performance of groundwater levels is a good measure to assess the performance of the parameter set assigned to the cell the borehole is located in as well as those within proximity to it. This is demonstrated in the results in Figure 7.4 where varied performance in the models is shown for the six different models. As soil and geological variability is up-scaled to a 1km × 1km grid, the work has assumed aggregated parameters. For example, (Kabat et al., 1997) describes aggregated soil parameters where the scale of grid cells results in soil parameters being represented by several soil types. In reality, heterogeneity of soil characteristics exist at microscopic, plot and field scales (10^{-2} – 10^{3} m) (Russo and Bresler, 1980), and so the difference in model performance at different sites may therefore be due to these aggregated parameter values being more representative of certain cells over others.

The results of this work therefore highlight the benefit and necessity for continual assessment at internal sites within the catchment, especially if the catchment is characteristically heterogeneous, such as the complex geology in the Tern catchment.

7.4. Summary of research questions and assessment of the main thesis aim

The main aim of the thesis has been to assess whether a certain framework (spatial representation, calibration method, and parameterisation) would result in a model that could be classified as better than the others according to the results of differing performance criteria as assessed at a range of flow gauges and groundwater level boreholes. The following summaries can be drawn for each objective:

Model spatial complexity – at the scale studied (a medium sized catchment) similar results were demonstrated for both complexities of models with similar statistics. The homogenous model, although a gross average of catchment characteristics, does not introduce human bias and uncertainty that is added by the spatial distributed model and the difficulty in finding representative parameter values for parameters that represent such a large area (1km). It was noted that at smaller scales, where data availability is

often better, and where studies specifically need a distributed model then such a model may be justified. In other cases, the difficulty in time required for no substantial improvement in results may highlight that a more homogenous model may be better.

Model calibration method – It was noted that different calibrations resulted in varying performances in different parts of the model. This was because different measures and statistics were given different emphases in each type of calibration. It is suggested that no model is statistically better than any other, but that for confidence in results, if developing a distributed model, it should be automatically calibrated using a range of measures and sites, so as to test the internal operation of the model. This was confirmed with the performance at groundwater level validation sites confirmed that for protocols 5 and 6 (multi-criteria automatic calibrations) that the mean validation scores were higher than the calibration mean score. In addition, for protocol 5 (homogenous multi-criteria automatic calibration) there was an increase from 17.1 to 17.2 from the overall mean calibration score to the mean validation score. This is in contrast to all the other protocols indicating an overall decrease from the calibration to the validation sites.

Parameterisation and equifinality of the parameter space – Although only a small area of research within this thesis which was addressed in Chapters 5 and 6, model results confirmed the existence of parameter equifinality – that more than one parameter set can result in output results which are statistically very similar. This research has built upon parameter equifinality as a result of calibrating a model using one method, and shown the existence of equifinality within the modelling protocol.

Issues of performance statistics - The differences discussed between quantitative statistics in not always suggesting the same levels of performance for a given model/output measure, suggests a need to be cautious when interpreting model performance from statistics alone. Despite appearing lengthy when fully documented as in this thesis, it also highlights the need to continually assess model performance using a range of both quantitative and qualitative methods including detailed inspection of plots of results against observed data internally within the catchment.

Internal measures of performance – Aside from differing levels of statistical performance shown at different sites, the key results of this research theme have been in

demonstrating the scaling effects apparent within the internal operation of the models. Most noticeable in the general poor performance at internal gauging stations such as Bailey Brook (calibration site) and Potford Brook (validation site), when compared to performance demonstrated at the basin outlet. In contrast, the simulated river flow regimes at the basin outlet for the six models are comparable. Scaling issues were shown, where results suggested catchment wide and grid scale parameterisations were more sensitive within smaller sub-basins, whereas at the outlet the effect of 'averaging' resulted in a more homogenous response of the models. Varied performance of model results at a range of groundwater level boreholes was also shown, with notable poor performance displayed close to the catchment boundary. Assessing performance at a range of different calibration and validation sites also confirmed internal ability of the models for sites at which they were not calibrated. For example, at Heathcote (Table 7.7) the mean score of 18 is comparable to the 18.8 scores at Edgmond and Longdon which were calibration sites.

It has been demonstrated that uncertainty relating to each of the research questions is too great to merit any one protocol being classified as better than the rest. In addition to the discussion detailed in this chapter, Table 7.7 provided a means of quantifying an evaluation of statistical model performance. Different models could be classified as performing generally better for river flows (the homogenous model automatically calibrated at the outlet) and groundwater levels (the homogenous model calibrated manually). As the aim has not been to prejudice one measure over another, qualitatively or statistically, none of the models can statistically be considered as better performing than the rest. This result is based on a conclusion that must be drawn from Table 7.7, that very often the ranges between the scores for the different protocols are quite minimal. Despite protocols 1 and 3 being ranked joint highest (score of 17.8), the lowest rated model (protocol 4) has an overall performance mean score of 16.8. All six protocols therefore demonstrate similar statistical ability.

7.5. Summary

This chapter has brought together the modelling results from chapters 5 and 6 to discuss key issues that comprised the research questions detailed in Chapter 1. The results that compared the six optimally calibrated models were presented and reference to them

made throughout discussions of each of the individual research questions (1-5) that this thesis assesses. A broad conclusion of each research theme was then made to support an answer to the main aim of the research; a decision that none of the methods or models (the modelling protocol) could justifiably be considered as better than the rest. In given respects, each of the different protocols were shown to have strengths and weakness, with the conclusion made that the purpose of any given research would need to determine the protocol/method a model is spatially conceptualised and calibrated.

Given the results of this chapter, the penultimate chapter of this thesis now follows to apply this result to a common application in hydrological modelling: assessing the impacts of climate change on water resources. As no firm decision could be made about a best performing model, rather than using one model to assess the impacts of climate change, the six calibrated models are used to assess uncertainty on water resources, and address what the implications of the uncertainty that was derived for models in the observed period, might mean in a wider sense relevant to applied studies undertaken in hydrological modelling.

Chapter 8

Climate change uncertainty analysis

8.1. Introduction

With uncertainty in model simulations from the six differently set-up and calibrated models of the present period (2000-2003) addressed in the previous chapter, this chapter continues the theme of model uncertainty with regard to a common application in hydrological modelling of climate change impact assessment on water resources. The standard approach to assess the impact of climate change on water resources outlined by Arnell and Reynard (2000) involves:

1. Definition, calibration and validation of a hydrological model using current climate data (the baseline scenario);
2. Definition of climate change scenarios and perturbation of the input climate data used in the original model;
3. Hydrological model simulation using the new input climate data and comparison of results with those of the baseline scenario.

As step one has already been completed within previous chapters, this chapter addresses steps 2 and 3. Section 8.2 (Step 2 above) provides a brief overview of modelling climate change impacts, and includes an analysis of the input climate data for the Tern catchment that is then used in Section 8.3 (Step 3 above). The new input climate data are then used to drive the six calibrated hydrological models and projected changes to Tern river flow and groundwater levels which are then discussed. The aim is not only to quantify potential climate change impacts in the Tern catchment, but in addition to make comparisons between the six hydrological models to assess how uncertainty between different models compares to uncertainty in the climate change signal for different scenarios. The research questions addressed in this chapter refer back to questions 6-8 in Chapter 1:

- What are the projected impacts of climate change simulated for the Tern catchment, using UKCIP02 data, for both river flows and groundwater levels at a range of locations?

- What is the uncertainty and range of simulated river flows and groundwater levels between the six calibrated hydrological models for different climate change scenarios and two time-slices (2050s and 2080s)?

- What is the magnitude of the intra-model uncertainty compared to the projected impacts of climate change, and how important is intra-model variability when using the models to simulate the potential effects of climate change?

8.2. Modelling climate change

With the well documented changes that have been occurring to global climate, such as the increase to global surface temperatures at 0.76 ^{o}C ± 0.19 ^{o}C from the last half of the 19th Century to the 2001-2005 period (Jenkins et al 2008), a substantial body of research is available on the impacts of changing climates on water resources both globally and at more regional scales (e.g. Arnell, 1999a; Arnell 1999b; Arnell, 2003; Sefton and Boorman, 1997; Thodson, 2007). In the temperate UK climate, warming temperatures are typically projected to result in increased evapotranspiration, and coupled with seasonal changes in the volume and distribution in precipitation, there are likely to be alterations from the current hydrological regimes of river flow and groundwater levels (Arnell and Reynard, 1996; IPCC, 2001).

8.2.1. Climate models and climate change scenarios

In order to undertake climate change impact studies on water resources, projections of changes to climate indices such as temperature, evapotranspiration and precipitation for future time periods are required as inputs into hydrological models. These projections of climate are typically derived from perturbed Coupled Ocean-Atmosphere General Circulation Models (OAGCMs) that are complex, physically based mathematical models that seek to simulate climate using a three dimensional Cartesian grid system that covers all or parts of the globe. A variety of OAGCMs have been developed that simulate the global circulation, typically at spatial resolutions such as 2.5º by 3.8º as represented in the UK's Hadcm3 model (Hulme et al., 2002; Jenkins et al, 2008).

These general circulation models describe the main dynamic and physical processes, their interactions and feedbacks (Ruosteenoja et al., 2003) and in general simulate the existing global climate (1961-1990) relatively well. Despite this, regional and sub-continental scales are often badly reproduced largely due to their large grid scale as well as poor representation of climate variables such as cloud formation, sea ice and surface-atmosphere interaction (Arnell and Reynard, 1996).

In the UK, a series of studies have been undertaken by the UK Climate Impacts Programme (UKCIP) in order to address the scaling issues. Regional Climate Models (RCMs) such as HadAM3H (resolution 120km) and HadRM3 (resolution 50km) have been developed to cover smaller areas such as Europe at a higher resolution, thus improving simulation of climate variables and features (Arnell and Reynard, 1996). These models use the larger GCM outputs as input boundary data, and thus are still linked to larger scale global circulation (Hulme et al., 2002).

So that climate change projections can be made in the UK, the regional climate circulation model HadRM3, for example, is then perturbed and forced with different climate change scenario data. These are based on a range of global emission scenarios such as those in Table 8.1. The Special Report on Emission Scenarios (SRES) define a range of possible future changes to greenhouse gases, the status of the economy and environment that feasibly could occur which range from the Low B1 scenario of efficient technologies and sustainable development, to the high A1F1 scenario

dominated by use of fossil fuels. Table 8.1 also shows how the four SRES scenarios that are often used in large scale OAGCMs are linked to similar higher resolution HadRM3 scenarios (Low (L), Medium-Low (ML), Medium-High (MH) and High (H)) . The scenario data are then used as inputs to re-simulate the regional circulation model for future time slices. The UKCIP02 data are available for three time slices for each scenario – the 2020s, the 2050s (50) and 2080s (80), relative to the baseline period 1961-1990. The simulated future climates can then be compared to the simulated current climate to assess potential changes in climate (Arnell and Reynard, 1996).

Table 8.1. Summary of UKCIP 02 scenarios and relation to SRES emission scenarios (adapted from UKCIP 01 report)

SRES Emission Scenario	**HadRM3 UKCIP02 Scenarios**	**Description**
B1	Low (L)	Clean and efficient technologies, reduction in material use, global solutions to economic social and environmental sustainability, improved equity, population peaks mid century.
B2	Medium Low (ML)	Local solutions to sustainability, continuously increasing population at a lower rate than A2, Less rapid technological change than in B1 and A1
A2	Medium High (MH)	Self-reliance, preservation of local identities, continually increasing population, economic growth on regional scales
A1F1	High (H)	V. Rapid economic growth, population peaks mid-century, social cultural and economic convergence among regions, market mechanisms dominate, reliance on fossil fuels

It is important to note that using climate models and future change scenarios as described here and subsequently employed in this thesis, is not the only method used to generate climate change data for climate change impact assessments. Simpler methods of assuming set rates of change for climate indices (for example, a $1^{o}C$ or $2^{o}C$ increase in temperature, or a 5 or 10% increase or decrease in precipitation) can also be used. For example Flickin et al., (2009) use this method in the SWAT model for the 14 983 km^2 San Joaquin catchment, California. Although these are crude climate change scenarios, they have been successfully used to assess the impact of overall changes in climate, especially with hydrological models operating on a monthly time step.

There are also more complex new generations of projected climate change data such as UKCP09 (released summer 2009 for the UK) that include probabilistic methodologies

that seek to further address the issues of uncertainty (Murphy et al., 2009). Ideally these data would have been used in this thesis but due to delay with their release, the timescale of this project prohibited its use.

A large quantity of research has been published that use different climate change scenarios derived from different OAGCM models that have been driven through many different hydrological models. Although these studies are usually not comparable because of the different models and scenarios used, they do show that climate change impacts on water resources such as river flows (Thodsen, 2007; Mernild et al, 2008; Steele-Dunne et al, 2008) or groundwater levels (Christensen et al., 2004; Fowler et al., 2003; Fowler et al., 2007; Thompson et al., 2009) will be large.

More recently, large scale projects such as QUEST-GSI (quest.bris.ac.uk) or HYACINTS (hyacints.dk) have been undertaken that use an established methodology to compare the impacts of climate change in different catchments. In addition, further awareness is being attributed to the uncertainty associated with results. For example, the QUEST-GSI project seeks to better quantify the impacts of climate change in a consistent way; partly by using outputs from the same seven general circulation models in different catchments including the Mekong, tributaries of the Nile (Kingston and Taylor, 2010), Yangtze and Yellow, Okavango, as well as a tributary of the Mackenzie. Results are then being compared to a larger scale global hydrological model to assess uncertainties of scale and whether the global and catchment models are comparable, and if so, by using a global scale model the impacts of climate change can be calculated in data sparse regions.

Despite recent research that addresses model uncertainty more comprehensively, the majority of climate change impact assessment work being undertaken usually does not assess uncertainty in results that may arise from uncertainty within the hydrological modelling protocol such as different methods of model calibration or model spatial representation, but from different climate model outputs and different future scenarios.

Wilby (2005) acknowledged some of the uncertainties in water resource projections by assessing the choice of calibration period, model structure and non-uniqueness of model parameter sets using the conceptual water balance model CATCHMOD for the River

Thames at Kingston. Additionally, Jiang et al (2007) sought to compare climate change impact results of six different hydrological models in the Donjiang Basin, South China. Here, large differences were seen between hydrological outputs of the six models, that, during the calibration and testing phases showed very similar regimes. In a similar method to Jiang et al (2007), the aim here is to compare the outputs of river flow and groundwater levels from different hydrological models, the six models that have already been documented in this thesis.

8.2.2. Projected climate change in the Tern Catchment

This section details the specific climate changes that are projected for the Tern catchment from UKCIP02 data (Hulme et al., 2002) that uses the 50km resolution HadRM3 regional climate model described previously. As shown in Figure 8.1, 91.7% of the Tern catchment is contained within HadRM3 Box 332, with only a small southern part of the catchment falling within Box 352. Given that Box 352 does not contain any of the principal stream network or locations of rain gauges, the projected changes to precipitation and temperature derived for Box 332 are used to perturb the new input climate data for the hydrological models.

Projected changes in monthly precipitation (%) and minimum and maximum temperature (^{o}C) were acquired from the UKCIP data distribution centre for the 2050s and 2080s time-slices for the four climate change scenarios (L, ML, MH, H) (Table 8.2).

As shown in Table 8.2 and Figure 8.2, precipitation is expected to decrease in summer months (April-October) and increase in winter months (November-March) Table 8.2 summarises the percentage changes for each month, winter and summer season, as well as the annual percentage change. In all scenarios an overall annual decrease in precipitation is projected ranging from -5.38% for the 50L scenario to -14.81% for the 80H scenario. In the most extreme scenario 80H, annual precipitation is projected to decrease by up to 51.29%, whilst at the same time increases in both minimum and maximum temperature (Figure 8.2) during the summer period of up to 4.16 ^{o}C (min temp) and 5.22 ^{o}C (max temp) suggest a significantly drier summer season, as evapotranspiration increases are linked to higher temperatures.

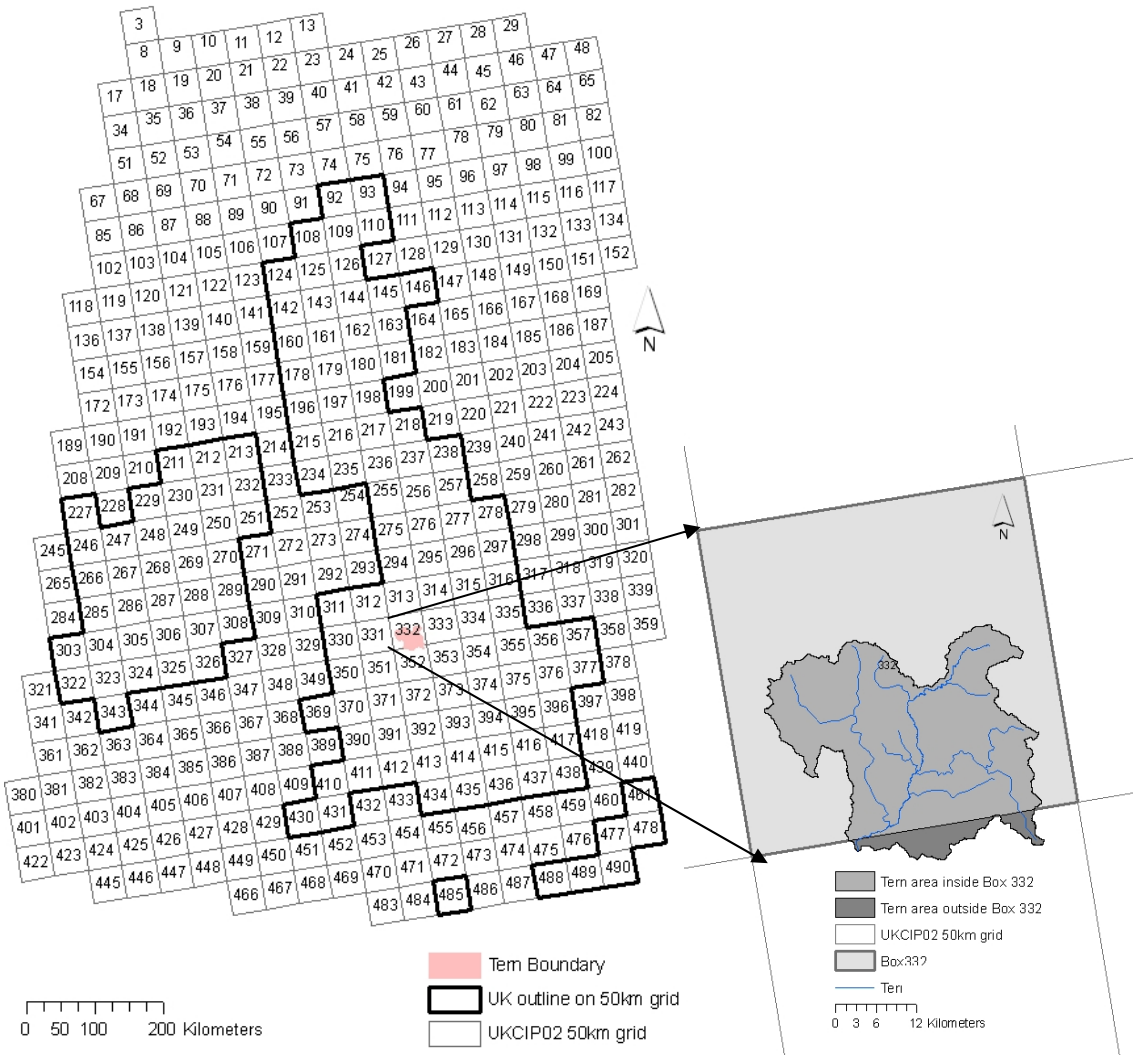


Figure 8.1. UKCIP02 50km^2 grid of the UK highlighting box 332 in which 91.7% of the Tern catchment is located

Table 8.3 and Figure 8.3 show the perturbed climate data (mean precipitation, mean PET and net precipitation) for the Tern catchment for each climate change scenario. The perturbed data have been derived using the commonly employed delta factor/change method (Prudhomme et al., 2002). The changes to precipitation for each of the eight climate change scenarios have been applied to the daily precipitation data from the baseline period (2000-2003), both for the catchment mean rainfall (for the homogenous model) and the eight individual gauged data that comprise the catchment mean (for the distributed model). The perturbed PET data were derived by re-evaluating the Hargreaves-Samani Eto calculation that was described in Section 3.6.2 using the projected maximum and minimum temperature data for each climate change scenario.

Although not a large problem, it is acknowledged that the baseline period in this study (2000-2003) is outside of the baseline period for UKCIP02 data (1961-1990). As noted by Thompson et al., (2009) the projected meteorological data are likely to be slightly over-estimated when compared to using a baseline period that was within the UKCIP02 baseline timescale. Additionally, the method of delta change perturbation does not account for any changes in the number of rain days or dry days that may potentially alter with climate change.

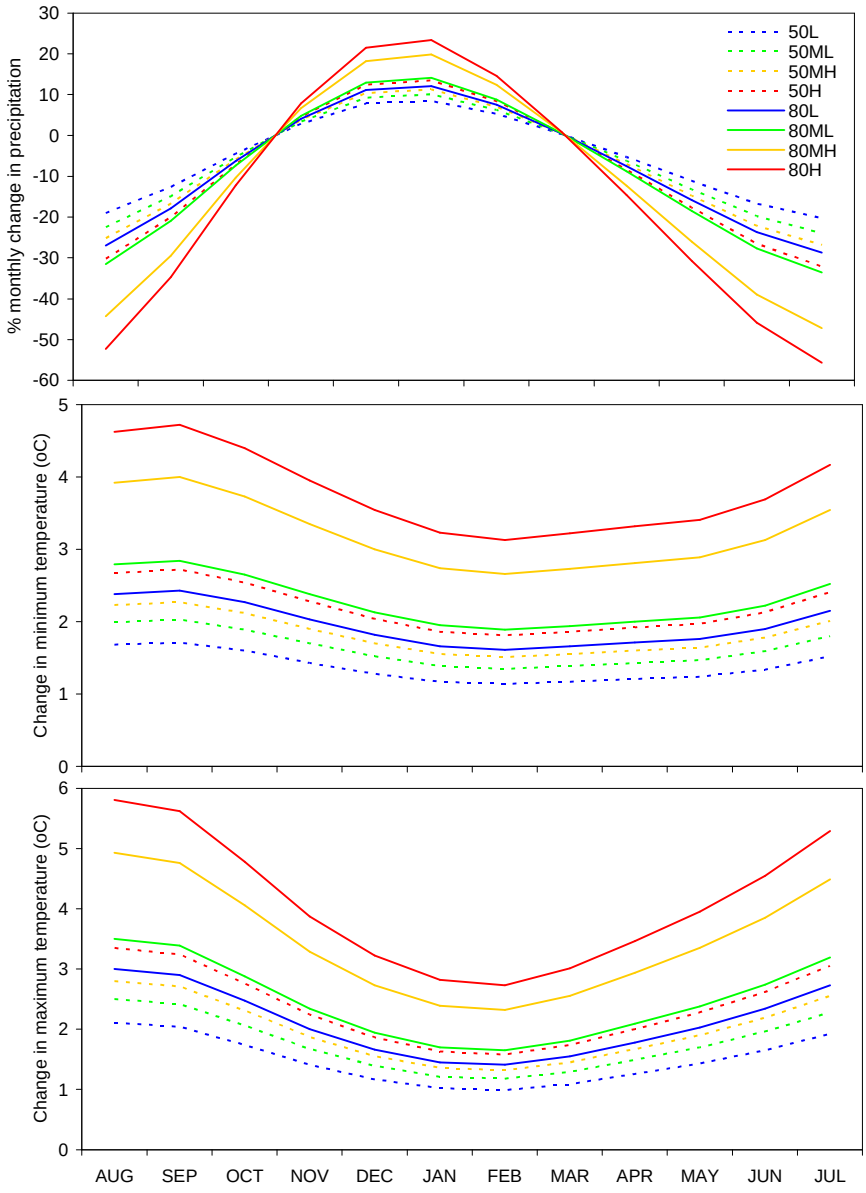


Figure 8.2 Monthly precipitation, minimum and maximum temperature projections for the eight UKCIP02 climate change scenarios

Table 8.2. Projected climate changes to precipitation (%) and temperature (min and max ^{o}C) in the West Midlands region (Box 332)

Variable	Period	Scenario	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	Winter DJF	Summer JJA	Annual
PPT	50s	L	-19.01	-12.61	-4.44	2.87	7.81	8.50	5.30	0.32	-5.28	-11.15	-16.68	-20.23	**7.20**	**-18.64**	**-5.38**
% change		ML	-22.51	-14.94	-5.26	3.40	9.25	10.07	6.27	0.38	-6.25	-13.21	-19.76	-23.96	**8.53**	**-22.08**	**-6.38**
		MH	-25.21	-16.73	-5.89	3.81	10.36	11.27	7.03	0.43	-7.00	-14.79	-22.13	-26.83	**9.55**	**-24.72**	**-7.14**
		H	-30.19	-20.04	-7.05	4.56	12.40	13.50	8.42	0.52	-8.39	-17.71	-26.50	-32.14	**11.44**	**-29.61**	**-8.55**
	80s	L	-26.96	-17.89	-6.30	4.07	11.08	12.05	7.51	0.46	-7.49	-15.82	-23.66	-28.70	**10.21**	**-26.44**	**-7.64**
		ML	-31.54	-20.94	-7.37	4.77	12.96	14.10	8.79	0.54	-8.76	-18.50	-27.69	-33.57	**11.95**	**-30.93**	**-8.93**
		MH	-44.35	-29.43	-10.36	6.70	18.22	19.83	12.36	0.76	-12.32	-26.02	-38.93	-47.21	**16.80**	**-43.49**	**-12.56**
		H	-52.30	-34.71	-12.22	7.90	21.49	23.38	14.58	0.89	-14.53	-30.68	-45.91	-55.67	**19.82**	**-51.29**	**-14.81**
Temp min	50s	L	1.68	1.71	1.60	1.43	1.28	1.17	1.14	1.17	1.21	1.24	1.34	1.52	**1.20**	**1.51**	**1.37**
^{o}C change		ML	1.99	2.03	1.89	1.70	1.52	1.39	1.35	1.39	1.43	1.47	1.59	1.80	**1.42**	**1.79**	**1.63**
		MH	2.23	2.27	2.12	1.90	1.70	1.55	1.51	1.55	1.60	1.64	1.78	2.01	**1.59**	**2.01**	**1.82**
		H	2.67	2.72	2.54	2.28	2.04	1.86	1.81	1.86	1.92	1.97	2.13	2.41	**1.90**	**2.40**	**2.18**
	80s	L	2.38	2.43	2.27	2.03	1.82	1.66	1.61	1.66	1.71	1.76	1.90	2.15	**1.70**	**2.15**	**1.95**
		ML	2.79	2.84	2.65	2.38	2.13	1.95	1.89	1.94	2.00	2.06	2.22	2.52	**1.99**	**2.51**	**2.28**
		MH	3.92	4.00	3.73	3.35	3.00	2.74	2.66	2.73	2.81	2.89	3.13	3.54	**2.80**	**3.53**	**3.21**
		H	4.62	4.72	4.40	3.95	3.54	3.23	3.13	3.22	3.32	3.41	3.69	4.17	**3.30**	**4.16**	**3.78**
Temp max	50s	L	2.11	2.04	1.74	1.41	1.17	1.02	0.99	1.09	1.26	1.43	1.65	1.92	**1.06**	**1.90**	**1.49**
^{o}C change		ML	2.50	2.42	2.06	1.67	1.39	1.21	1.18	1.29	1.49	1.70	1.96	2.28	**1.26**	**2.25**	**1.76**
		MH	2.80	2.71	2.31	1.87	1.55	1.36	1.32	1.45	1.67	1.90	2.19	2.55	**1.41**	**2.51**	**1.97**
		H	3.35	3.24	2.76	2.24	1.86	1.63	1.58	1.74	2.00	2.28	2.62	3.05	**1.69**	**3.01**	**2.36**
	80s	L	3.00	2.90	2.47	2.00	1.66	1.45	1.41	1.55	1.78	2.03	2.34	2.73	**1.51**	**2.69**	**2.11**
		ML	3.50	3.39	2.88	2.34	1.94	1.70	1.65	1.81	2.09	2.38	2.74	3.19	**1.76**	**3.15**	**2.47**
		MH	4.93	4.76	4.06	3.29	2.73	2.39	2.32	2.55	2.94	3.35	3.85	4.49	**2.48**	**4.42**	**3.47**
		H	5.81	5.62	4.78	3.87	3.22	2.82	2.73	3.01	3.46	3.95	4.55	5.29	**2.92**	**5.22**	**4.09**

Table 8.3 highlights the total annual precipitation for the baseline period was 761.3 mm and indicates an overall decreasing trend in precipitation for future climate change scenarios. The 50L scenario projects a 5.5% decrease in the catchment, this rising to a 15% decrease in the 80H scenario. Further analyses of changes to projected Tern precipitation are made by assessing changes to seasonality (winter and summer). For the present baseline period there is little difference between total winter (DJF) and summer precipitation (JJA) (177.5 mm and 177.3 mm respectively).

As shown in Table 8.3 and Figure 8.3 seasonal differences in precipitation are projected to become more marked with projected climate change. The summer precipitation, 177.3 mm for the baseline period reduces to 143.9 mm for the 50L, and to 85.3 mm for the 80H scenarios. In winter, the baseline precipitation is 177.5 mm. For the 50L scenario this increases to 190.0 mm, and for the 80H scenario to 212.3 mm.

Table 8.3 and Figure 8.3 also show the projected PET for the Tern catchment for each scenario. A trend of increasing annual PET is shown with the baseline of 647.2 mm increasing for the 50L scenario to 688.0 mm, and for the 80H scenario a further increase to 763.6 mm. Smaller increases in PET are seen during winter with an increase of 10.7% for the 80H scenario, compared to a 20.2% summer increase.

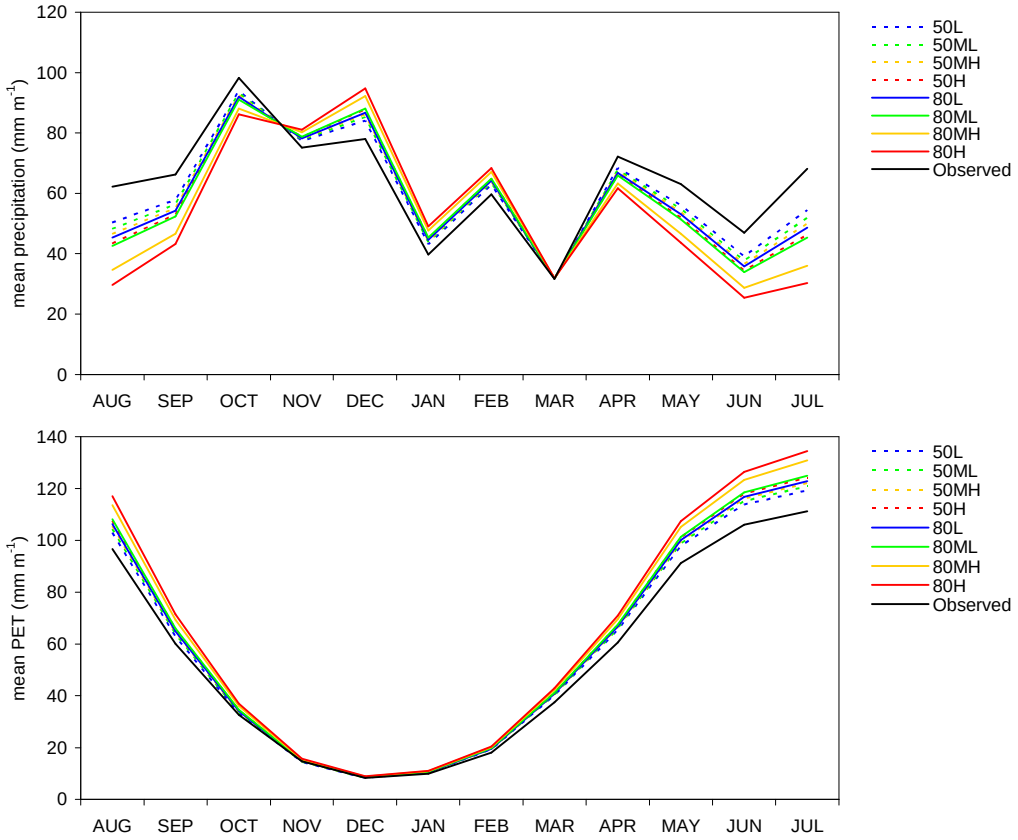


Figure 8.3. Mean monthly precipitation and PET for eight climate change scenarios in the Tern Catchment

8.3. Hydrological modelling of uncertainty of impacts of climate change in the Tern Catchment

This section addresses the third task in climate change impact assessment; the use of the hydrological models to simulate the impacts of climate change. The new input climate data are used and comparison of results are made to the baseline calibrations using the original climate data (Arnell and Reynard, 1996).

Table 8.3. Average monthly, and total seasonal and annual precipitation, potential evapotranspiration and net precipitation for each climate change scenario (mm)

Variable	Period	Scenario	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	Total Winter DJF	Total Summer JJA	Total Annual
PPT		Baseline	**62.2**	**66.2**	**98.2**	**75.2**	**78.0**	**39.8**	**59.7**	**31.6**	**72.2**	**63.1**	**46.9**	**68.2**	**177.5**	**177.3**	**761.3**
	50s	L	50.4	57.8	93.9	77.3	84.1	43.1	62.8	31.7	68.4	56.1	39.1	54.4	190.0	143.9	719.1
		ML	48.2	56.3	93.1	77.7	85.2	43.8	63.4	31.7	67.7	54.8	37.7	51.9	192.4	137.8	711.5
		MH	46.5	55.1	92.4	78.1	86.1	44.2	63.9	31.8	67.1	53.8	36.5	49.9	194.2	132.9	705.4
		H	43.4	52.9	91.3	78.6	87.7	45.1	64.7	31.8	66.1	51.9	34.5	46.3	197.5	124.2	694.3
	80s	L	45.4	54.3	92.0	78.2	86.7	44.6	64.2	31.8	66.8	53.1	35.8	48.6	195.5	129.8	701.5
		ML	42.6	52.3	91.0	78.8	88.1	45.4	64.9	31.8	65.9	51.4	33.9	45.3	198.4	121.8	691.4
		MH	34.6	46.7	88.0	80.2	92.2	47.7	67.1	31.9	63.3	46.7	28.7	36.0	207.0	99.3	663.1
		H	29.7	43.2	86.2	81.1	94.8	49.1	68.4	31.9	61.7	43.7	25.4	30.2	212.3	85.3	645.4
PET		Baseline	**96.6**	**60.1**	**32.6**	**14.7**	**8.4**	**10.0**	**18.1**	**37.5**	**60.7**	**91.2**	**106.1**	**111.2**	**36.5**	**313.9**	**647.2**
	50s	L	102.9	62.9	33.1	14.4	8.4	10.4	19.3	40.2	65.4	97.9	113.8	119.3	38.1	336.0	688.0
		ML	104.4	63.8	33.6	14.6	8.4	10.5	19.4	40.4	66.0	98.9	115.2	120.9	38.3	340.5	696.1
		MH	105.5	64.5	33.9	14.7	8.5	10.6	19.5	40.7	66.4	99.7	116.1	122.1	38.6	343.7	702.2
		H	107.6	65.7	34.4	14.9	8.5	10.7	19.7	41.1	67.3	101.1	118.0	124.3	38.9	349.9	713.3
	80s	L	106.3	64.9	34.1	14.7	8.5	10.6	19.6	40.8	66.7	100.1	116.8	122.9	38.7	346.0	706.0
		ML	108.1	66.1	34.6	14.9	8.6	10.7	19.7	41.2	67.5	101.4	118.5	124.9	39.0	351.5	716.2
		MH	113.6	69.4	36.1	15.4	8.8	10.9	20.1	42.3	69.6	105.1	123.3	130.8	39.8	367.7	745.4
		H	117.0	71.5	37.0	15.7	8.9	11.0	20.4	42.9	70.9	107.4	126.4	134.5	40.3	377.9	763.6

Using the six 'optimally' calibrated hydrological models, the perturbed daily precipitation and evapotranspiration data described in the previous section were input to each model and the simulations re-run. In total, each of the six models have been perturbed with the four scenarios (L, ML, MH, H) and for the two separate time slices (2050s and 2080s) which provided a total of 48 simulations. For the three distributed models with different calibrations, the precipitation data have been perturbed for each of the eight gauges from which data is then distributed using Thiessen polygons. For the homogenous model, the catchment mean rainfall from these eight gauges has been perturbed. All the calibration parameters for each of the models have been kept the same as described in chapters 4, 5 and 6, with only the input climate data altered.

Rather than only assessing climate change impacts at the basin outlet, this work assesses the projected differences and uncertainties at the four river flow gauging stations assessed throughout the thesis. Additionally, the impacts of climate change are also assessed at the same three groundwater level boreholes also assessed previously in the thesis.

The objective of this work is not to only assess any impacts of climate change on water resources in the catchment, objective one (Section 8.3.1), but the aim is also to assess if all six models that have been constructed differently, and calibrated differently, produce similar projections, both in magnitude and seasonal patterns, and to assess how different internal gauges and boreholes within the model respond to the new climate change scenario data, objective two (Section 8.3.2). The last objective is to quantify the magnitude of these uncertainties and climate change projections relative to each other (Section 8.3.3).

8.3.1. Assessment of climate change impacts in the Tern catchment

In order to assess the impact of climate change on water resources in the Tern catchment (the first of the three objectives of this chapter), results of daily river flows at the basin outlet are shown for the six models in Figure 8.4 and 8.5 for typically wet (2000-2001), and dry years (2002-2003) respectively. For each model, results for all eight climate change scenarios are shown.

It is apparent that there is a general drying trend in river flows at the basin outlet, Walcot. This is shown in the results of the majority of the six models. Decreases in

flows are shown throughout the year, with the exception of increases to individual winter peak events (November 2000, Figure 8.4, and December 2002 – January 2003, Figure 8.5). In both the drying trend, and the increases in winter peak flows for individual events, the 80H scenario shows the most extreme impact. There appears to be a progressive trend in drying for the emission scenarios (low to high) for each time slice (2050s and 2080s).

The peaks that occur outside of the winter season such as those in September and October 2000, and June 2001 (Figure 8.4), and July 2003 (Figure 8.5) do not increase under climate change scenarios. In all cases the simulated peaks are much lower under the climate change scenarios due to the combination of increased temperatures (enhancing Eto) and decreased rainfall during these months. Other studies in England have also noted increased seasonal variability, with more runoff in winter and a decrease during summer months (Arnell, 1998; Pilling and Jones, 1999).

Despite these general patterns which are shown by the majority of the six models, inconsistencies in the simulated impacts of climate change are also shown between models. For example, between December 2000 and March 2001 (Figure 8.4), the homogenous manually calibrated model (solid blue line) shows decreases in flows almost consistently throughout the year with the 80H scenario exhibiting the largest reduction. In contrast, the distributed models (dashed lines) show very little variation between climate change scenarios for the same period. Likewise, this pattern is repeated during the 2002-2003 year in Figure 8.5.

Deviations of mean monthly river flow from the baseline calibrated period are shown in Figure 8.6. This figure shows more clearly the impacts of climate change for the two extreme scenarios of the two time slices (50L and 80H), at both the basin outlet as well as internal gauging stations to assess whether the impacts of climate change are consistent throughout the catchment. The drying trend in both the 50L and 80H scenarios is confirmed with the deviations from the baseline predominantly plotted beneath the line where y=0, the baseline calibration. It is apparent that the 80H scenario results in the largest decrease in river flows, with MDF from the six models reduced from 6.90 m^3s^{-1} to 4.82 m^3s^{-1} at the basin outlet (shown later in Table 8.6 for baseline and the 80H scenario).

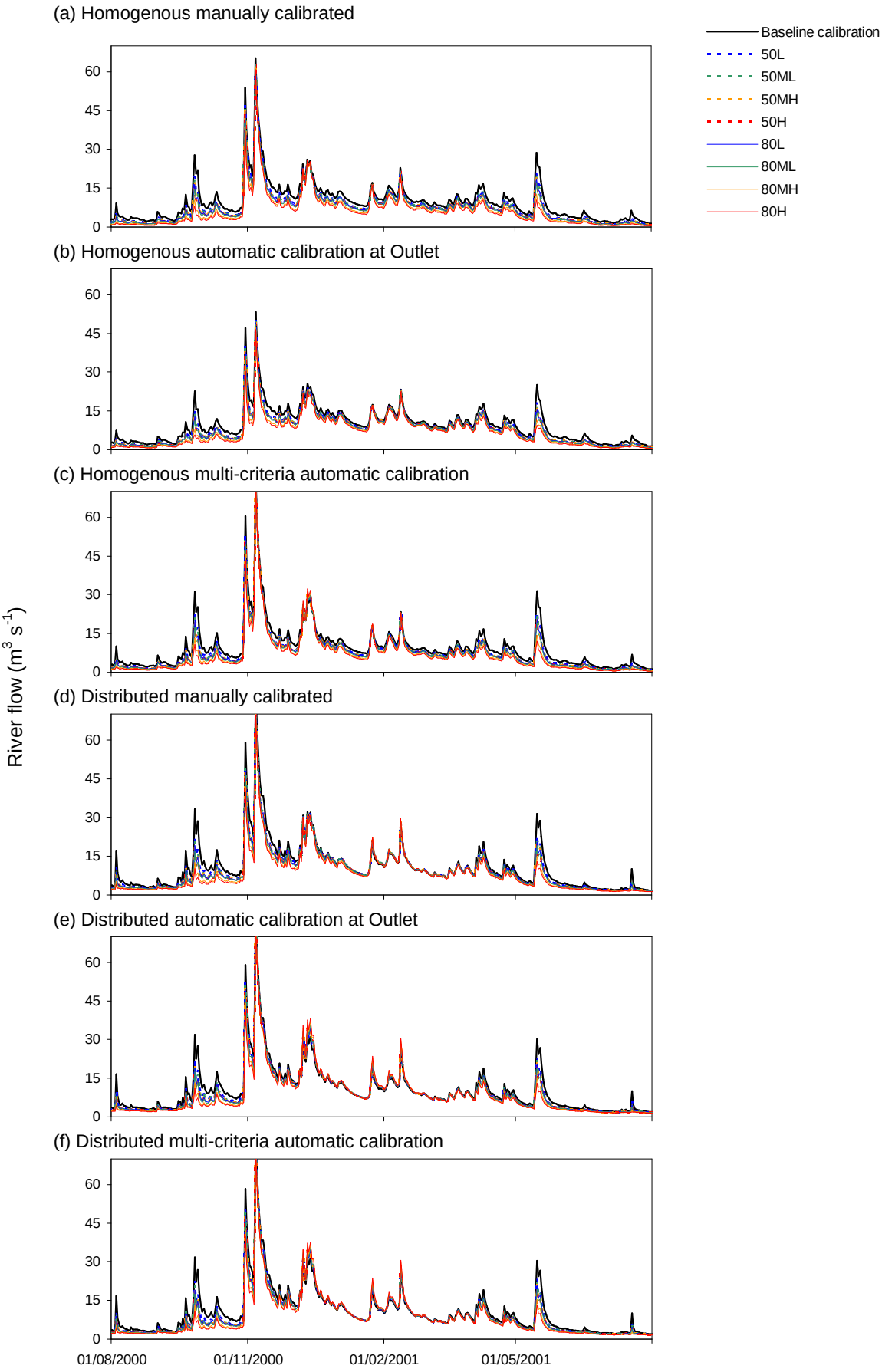


Figure 8.4 Daily simulated river flows for the six hydrological models for the eight climate change scenarios for the characteristically wet 2000-2001 year

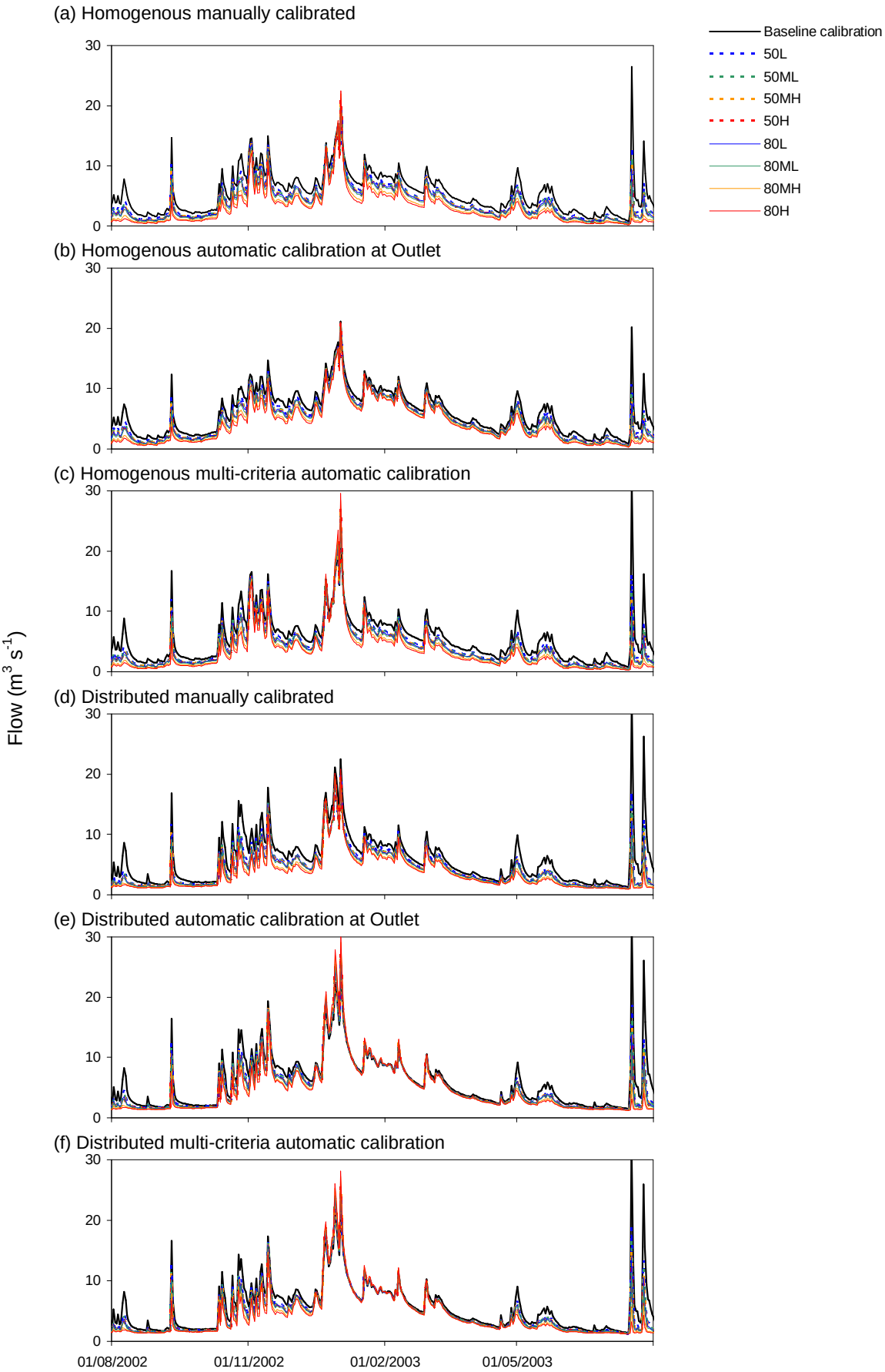


Figure 8.5 Daily simulated river flows for the six hydrological models for the eight climate change scenarios for the characteristically dry 2002-2003 year

Figure 8.6 also demonstrates the widely varying patterns between the six different models. Although the issue of intra-model uncertainty is discussed further in Section 8.3.2, two of the models (homogenous, manually calibrated (solid blue line), and the distributed automatically calibrated using the multi-criteria approach (dashed red line)), highlight the differences between the models in the simulated impacts of climate change. For example, the multi-criteria automatically calibrated distributed model indicates peak flows to increase during wet winter months, whereas the homogenous manually calibrated model indicates decreases in river flows throughout the year for both the 50L and 80H scenario. However, in general, similar climate change impacts for the two scenarios are shown at all the gauging stations, with a general overall decrease in flow volume. This is also shown in Table 8.6, which demonstrates that the MDF for the six models decreases in all cases for the 50L scenario, and also for the 80H scenario.

Figure 8.7 displays the mean monthly deviations of groundwater levels at three borehole locations (using the same method as Figure 8.6). The daily output plots for groundwater levels have not been included, as groundwater level variation on a daily scale is minimal in comparison to river flows. Figure 8.7 indicates a drying trend with general overall decreases in groundwater levels. Despite this, the three borehole locations indicate different magnitudes of declining groundwater levels. At Longdon, the levels are within -0.5m of the baseline for the 50L scenario and generally within -1.0m for the 80H scenario. In contrast, at the Gnosall borehole the model simulates levels to decline by up to 1.0m and 2.5m for the two scenarios, respectively. The differences between sites are likely due to the earlier described differing performances during the model calibration stages. For example, at Longdon, the calibrations were relatively good where as at Gnosall the poor calibrations and greater uncertainty may be causing further uncertainty in this impact assessment stage.

In addition to the differing projected impacts of climate change at different boreholes, the intra-model variation and uncertainty between models at each site (especially at Gnosall), highlights the issue that models poorly calibrated can result in widely differing results when using them to simulate impacts of catchment change, to the extent that they may be considered of no beneficial use at all. It is therefore difficult, apart from identifying a general drying trend, to further quantify the impacts of climate change on groundwater levels with any credibility.

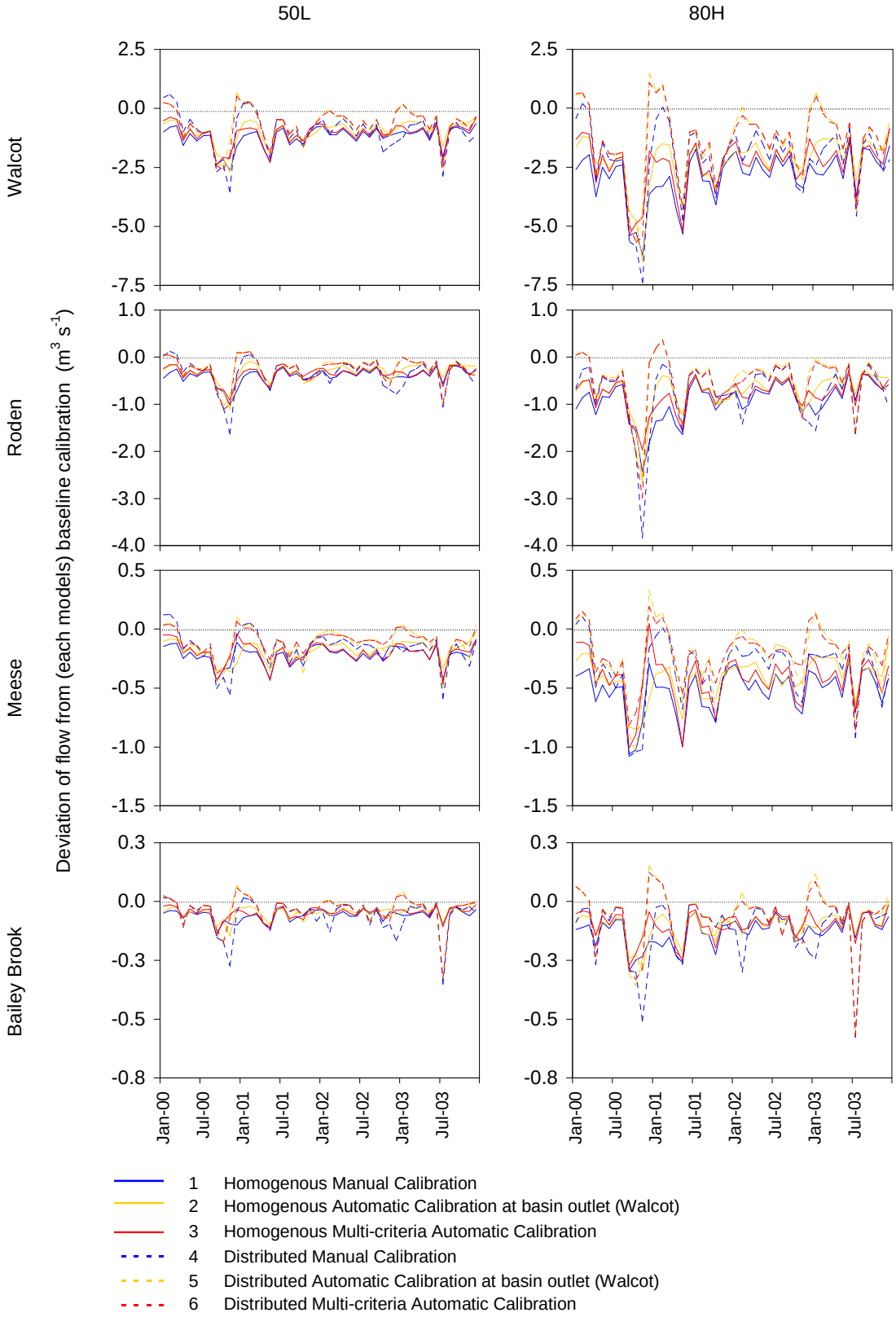


Figure 8.6. Deviations of monthly mean river flow of 50L and 80H climate change scenarios from the calibrated baseline for each of the six hydrological models, and at four gauging stations

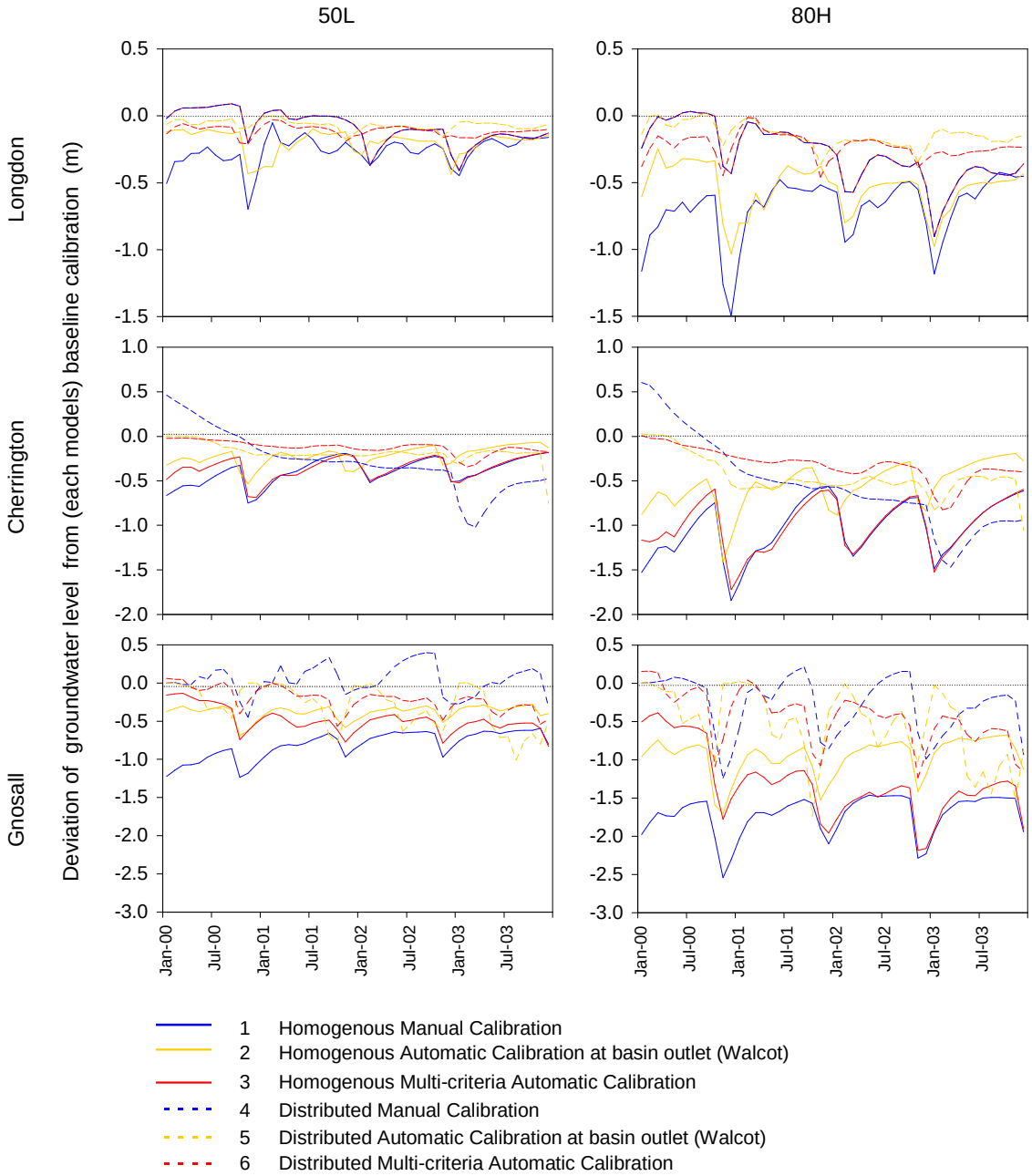


Figure 8.7. Deviations of monthly groundwater levels of 50L and 80H climate change scenarios from the calibrated baseline for each of the six hydrological models, and at three boreholes

8.3.2. Hydrological model uncertainty in simulating the impacts of climate change

Having briefly assessed the projected impacts of climate change on river flows and groundwater levels in the Tern catchment, this section focuses on further comparing model uncertainty between the six calibrated hydrological models. Regimes of river flow and groundwater levels for the 2000-2003 period (Figures 8.8 and 89), as well as tables showing the uncertainty between the six models (maximum – minimum) for the range of scenarios (Tables 8.4 and 8.5) are used to describe and assess model uncertainty.

Figure 8.8 shows the river flow regimes mean monthly flows for the baseline (calibrated period), 50L and 80H scenarios. The same y axis scales are used for each gauging station to enable comparisons. The regimes demonstrate the general drying trend at each gauging station with both the mean monthly summer low flows and winter peak flows substantially reduced for the 50L and 80H simulations. It is noted that earlier increases projected for individual winter peak river flow events are not shown at the mean monthly scale.

Aside from the general drying pattern, there is also an increase in the variation and uncertainty between the six models. For example, at Walcot, the 0.72 m^3s^{-1} (10.4% mean variation) range between models for the baseline period is increased to 1.93 m^3s^{-1} (or 40.1% variation between models) by the 80H scenario (Table 8.4). This pattern is also reproduced for the other internal flow gauging stations. Notably at Bailey Brook, the 0.16 m^3s^{-1} mean range between the six models for the 80H scenario translates to a 95.2% variation in simulated mean daily flow as a proportion of the baseline mean daily flow (Table 8.4).

In addition to the increasing uncertainty between the models, the regime shapes/seasonality are also simulated differently by the six models for the different climate change scenarios. Using the Roden as an example, and comparing the manually calibrated homogenous model (solid blue line) with the automatically multi-objective calibration for the distributed model (dashed red-line), it can be seen that the regime for the latter model shows a change in the peak flow month from November for the baseline period, to February for the 80H scenario. Comparatively, the homogenous manually calibrated model does not show this shift, but a regime which is similar to the baseline period is seen for the 80H scenario.

Table 8.4 shows that for each time slice (2050s and 2080s), the mean river flow of the six models increases with progressively higher emission scenarios. The trend is enhanced for the 2080s time slice compared to the 2050s, and indicates that intra model uncertainty is largest for the most extreme climate change scenario. The trend in increasing model uncertainty is shown annually and for both summer (June-August) and winter (December-February) periods.

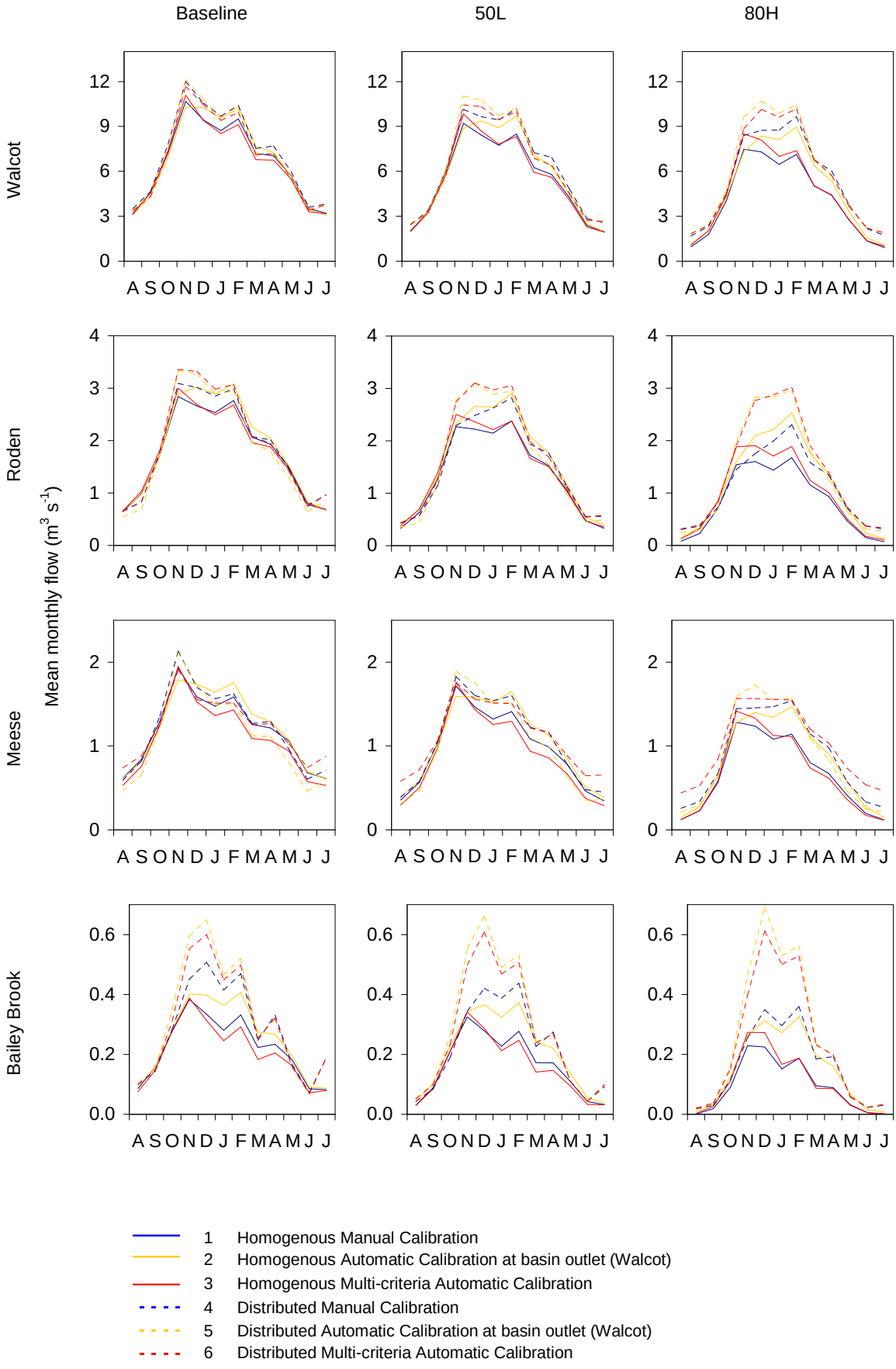


Figure 8.8. Simulated annual river flow regimes of the baseline period, 50L and 80H scenarios for the six models, at four gauging stations

Table 8.4. Monthly, seasonal and annual ranges in simulated river flows (m^3s^{-1}) for each climate change scenario, at four gauging stations, to compile hydrological model uncertainty (Calculated by monthly maximum – minimum for the range of six models)

		A	S	O	N	D	J	F	M	A	M	J	J	Summer	Winter	Annual
Baseline	Walcot	0.41	0.35	0.87	1.83	1.44	1.18	1.30	0.92	0.94	0.50	0.30	0.79	0.50	1.30	0.90
	Roden	0.12	0.34	0.10	0.51	0.66	0.49	0.40	0.30	0.26	0.24	0.20	0.29	0.20	0.52	0.33
	Meese	0.27	0.23	0.13	0.34	0.21	0.28	0.33	0.29	0.23	0.27	0.28	0.35	0.30	0.27	0.27
	BB	0.03	0.01	0.07	0.21	0.34	0.22	0.23	0.09	0.13	0.04	0.02	0.11	0.05	0.26	0.13
50L	Walcot	0.52	0.18	0.45	2.16	2.35	1.98	1.90	1.30	1.34	0.73	0.59	0.71	0.61	2.08	1.19
	Roden	0.11	0.25	0.24	0.51	0.88	0.83	0.67	0.42	0.26	0.19	0.10	0.25	0.15	0.79	0.39
	Meese	0.29	0.25	0.10	0.30	0.32	0.28	0.35	0.33	0.30	0.25	0.30	0.36	0.32	0.32	0.29
	BB	0.02	0.02	0.07	0.23	0.38	0.28	0.29	0.10	0.13	0.04	0.02	0.07	0.04	0.32	0.14
50ML	Walcot	0.59	0.22	0.43	2.20	2.52	2.20	2.10	1.39	1.40	0.78	0.64	0.75	0.66	2.27	1.27
	Roden	0.13	0.23	0.24	0.51	0.93	0.89	0.75	0.43	0.28	0.18	0.11	0.25	0.16	0.86	0.41
	Meese	0.30	0.25	0.12	0.29	0.34	0.29	0.35	0.33	0.32	0.25	0.30	0.37	0.32	0.33	0.29
	BB	0.02	0.02	0.07	0.23	0.39	0.29	0.30	0.11	0.13	0.04	0.02	0.06	0.03	0.33	0.14
50MH	Walcot	0.63	0.24	0.40	2.20	2.61	2.32	2.18	1.42	1.42	0.81	0.66	0.77	0.69	2.37	1.30
	Roden	0.14	0.21	0.23	0.53	0.96	0.93	0.79	0.44	0.28	0.17	0.12	0.25	0.17	0.89	0.42
	Meese	0.30	0.25	0.13	0.28	0.34	0.29	0.36	0.34	0.33	0.25	0.30	0.36	0.32	0.33	0.29
	BB	0.02	0.02	0.07	0.23	0.40	0.29	0.30	0.11	0.13	0.04	0.02	0.06	0.03	0.33	0.14
50H	Walcot	0.71	0.35	0.46	2.25	2.86	2.63	2.49	1.53	1.49	0.88	0.71	0.83	0.75	2.66	1.43
	Roden	0.17	0.18	0.22	0.55	1.04	1.03	0.91	0.45	0.31	0.17	0.14	0.25	0.18	0.99	0.45
	Meese	0.31	0.26	0.15	0.27	0.35	0.33	0.37	0.35	0.35	0.28	0.31	0.36	0.33	0.35	0.31
	BB	0.02	0.02	0.06	0.23	0.42	0.31	0.32	0.12	0.13	0.04	0.02	0.06	0.03	0.35	0.15
80L	Walcot	0.66	0.27	0.41	2.22	2.71	2.43	2.29	1.46	1.45	0.83	0.67	0.79	0.71	2.47	1.35
	Roden	0.15	0.20	0.23	0.54	0.99	0.96	0.83	0.44	0.29	0.16	0.12	0.25	0.17	0.93	0.43
	Meese	0.31	0.25	0.14	0.27	0.34	0.31	0.36	0.34	0.34	0.26	0.30	0.36	0.32	0.34	0.30
	BB	0.02	0.02	0.07	0.23	0.41	0.30	0.31	0.11	0.13	0.04	0.02	0.06	0.03	0.34	0.14
80ML	Walcot	0.73	0.39	0.48	2.22	2.92	2.72	2.57	1.56	1.50	0.89	0.73	0.85	0.77	2.74	1.46
	Roden	0.17	0.17	0.21	0.55	1.06	1.06	0.93	0.46	0.32	0.18	0.14	0.25	0.19	1.02	0.46
	Meese	0.32	0.26	0.16	0.22	0.35	0.35	0.37	0.36	0.36	0.29	0.32	0.36	0.33	0.36	0.31
	BB	0.02	0.02	0.06	0.23	0.42	0.31	0.32	0.12	0.12	0.04	0.02	0.05	0.03	0.35	0.15
80MH	Walcot	0.88	0.64	0.72	2.34	3.40	3.43	3.33	1.86	1.62	0.99	0.88	0.98	0.91	3.39	1.76
	Roden	0.21	0.11	0.17	0.57	1.17	1.30	1.19	0.64	0.39	0.22	0.20	0.26	0.22	1.22	0.54
	Meese	0.32	0.29	0.23	0.22	0.43	0.42	0.41	0.43	0.40	0.35	0.35	0.35	0.34	0.42	0.35
	BB	0.02	0.02	0.07	0.23	0.45	0.35	0.36	0.14	0.12	0.04	0.02	0.04	0.03	0.39	0.15
80H	Walcot	0.97	0.75	0.90	2.52	3.70	3.85	3.81	2.19	1.75	1.10	0.99	1.05	1.00	3.79	1.96
	Roden	0.24	0.16	0.15	0.55	1.22	1.44	1.34	0.75	0.45	0.27	0.23	0.26	0.24	1.33	0.59
	Meese	0.32	0.30	0.28	0.31	0.49	0.48	0.46	0.46	0.42	0.37	0.36	0.34	0.34	0.48	0.38
	BB	0.02	0.02	0.07	0.24	0.47	0.38	0.38	0.15	0.11	0.04	0.02	0.03	0.02	0.41	0.16

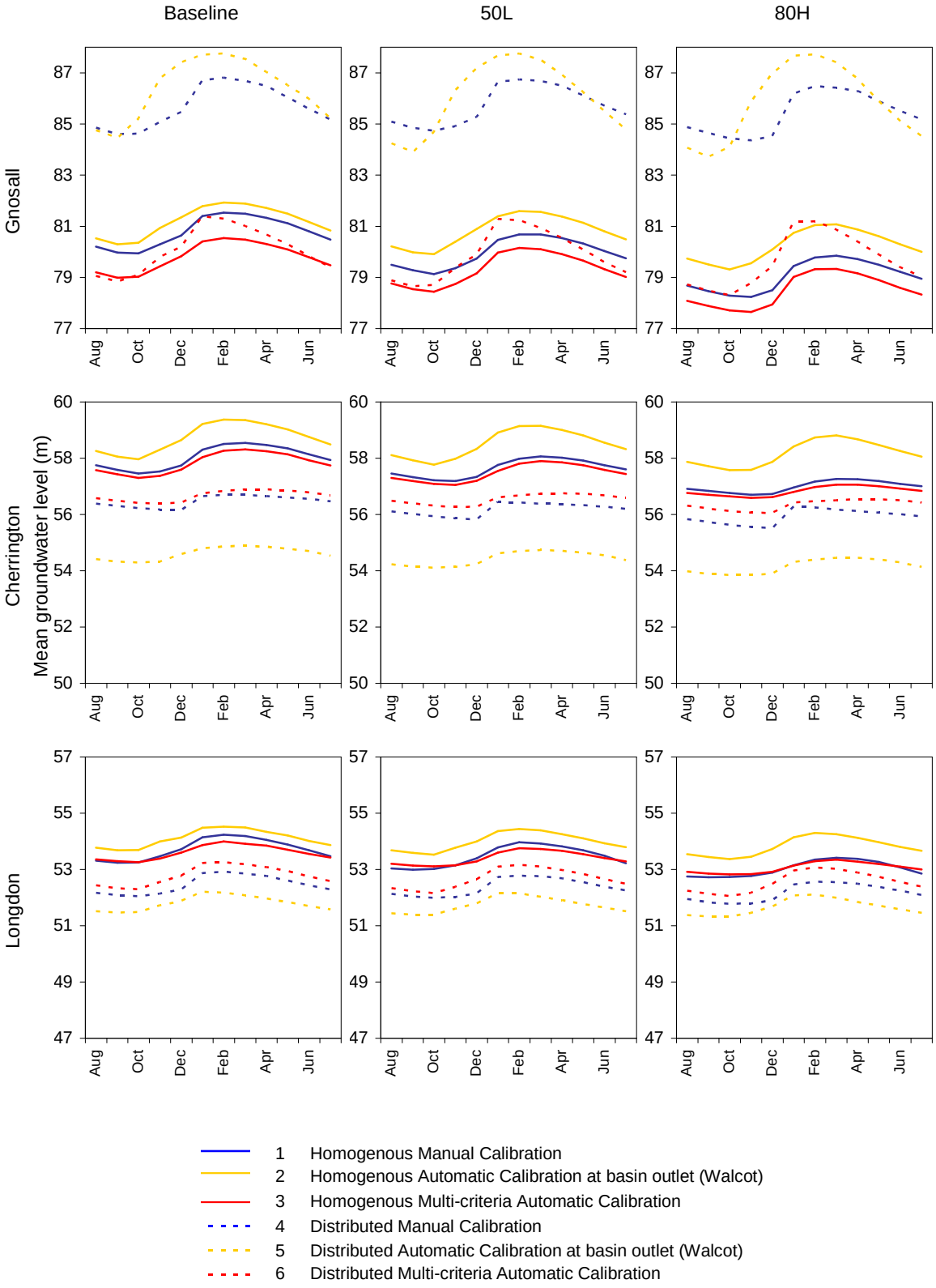


Figure 8.9. Simulated groundwater level regimes at three boreholes for the baseline period, 50L and 80H scenarios, compared for the six calibrated models

Table 8.5. Monthly, seasonal and annual ranges in simulated groundwater levels (m) for each climate change scenario, at three boreholes, to compile hydrological model uncertainty. (Calculated by monthly maximum – minimum for the range of six models)

		A	S	O	N	D	J	F	M	A	M	J	J	Summer	Winter	Annual
Baseline	Gnosall	5.79	5.76	6.18	7.34	7.58	7.30	7.22	7.06	6.72	6.41	6.19	5.83	5.94	7.37	6.62
	Cherrington	3.84	3.72	3.66	3.97	4.05	4.41	4.51	4.46	4.36	4.24	4.04	3.95	3.94	4.32	4.10
	Longdon	2.25	2.21	2.19	2.26	2.25	2.28	2.35	2.42	2.36	2.36	2.30	2.28	2.28	2.29	2.29
50L	Gnosall	6.33	6.30	6.29	7.58	8.02	7.71	7.60	7.39	7.01	6.59	6.39	6.35	6.36	7.78	6.96
	Cherrington	3.87	3.77	3.65	3.84	4.10	4.30	4.44	4.41	4.29	4.16	4.00	3.94	3.94	4.28	4.06
	Longdon	2.23	2.18	2.12	2.16	2.20	2.21	2.29	2.36	2.35	2.32	2.28	2.26	2.26	2.23	2.25
50ML	Gnosall	6.38	6.35	6.36	7.65	8.13	7.80	7.67	7.45	7.05	6.66	6.46	6.41	6.42	7.87	7.03
	Cherrington	3.87	3.78	3.65	3.82	4.07	4.27	4.43	4.40	4.28	4.14	3.99	3.93	3.93	4.26	4.05
	Longdon	2.21	2.17	2.11	2.13	2.17	2.18	2.27	2.34	2.33	2.30	2.26	2.25	2.24	2.21	2.23
50MH	Gnosall	6.41	6.38	6.39	7.69	8.20	7.86	7.73	7.49	7.09	6.66	6.48	6.44	6.44	7.93	7.07
	Cherrington	3.87	3.78	3.66	3.80	4.05	4.23	4.40	4.38	4.26	4.13	3.98	3.93	3.93	4.23	4.04
	Longdon	2.21	2.16	2.11	2.12	2.16	2.17	2.27	2.34	2.33	2.30	2.25	2.24	2.23	2.20	2.22
50H	Gnosall	6.52	6.49	6.51	7.77	8.33	8.00	7.85	7.59	7.16	6.68	6.61	6.56	6.56	8.06	7.17
	Cherrington	3.88	3.79	3.66	3.78	4.02	4.21	4.39	4.38	4.25	4.11	3.97	3.93	3.93	4.21	4.03
	Longdon	2.20	2.15	2.09	2.08	2.13	2.15	2.25	2.31	2.32	2.28	2.24	2.23	2.22	2.18	2.20
80L	Gnosall	6.48	6.45	6.47	7.71	8.25	7.90	7.78	7.53	7.12	6.68	6.55	6.51	6.51	7.98	7.12
	Cherrington	3.87	3.78	3.66	3.79	4.04	4.22	4.39	4.38	4.26	4.12	3.98	3.93	3.93	4.22	4.04
	Longdon	2.21	2.16	2.10	2.11	2.15	2.17	2.26	2.33	2.33	2.29	2.25	2.24	2.23	2.19	2.22
80ML	Gnosall	6.54	6.51	6.52	7.82	8.39	8.05	7.88	7.61	7.19	6.70	6.63	6.58	6.58	8.11	7.20
	Cherrington	3.89	3.80	3.67	3.78	4.03	4.21	4.40	4.38	4.26	4.12	3.98	3.94	3.94	4.21	4.04
	Longdon	2.20	2.14	2.09	2.08	2.13	2.15	2.25	2.31	2.32	2.28	2.24	2.23	2.22	2.18	2.20
80MH	Gnosall	6.68	6.65	6.63	8.03	8.74	8.42	8.21	7.90	7.44	6.89	6.80	6.73	6.74	8.46	7.43
	Cherrington	3.90	3.81	3.69	3.74	3.98	4.13	4.35	4.34	4.22	4.07	3.95	3.92	3.92	4.15	4.01
	Longdon	2.18	2.12	2.06	2.03	2.07	2.10	2.22	2.29	2.30	2.26	2.22	2.21	2.20	2.13	2.17
80H	Gnosall	6.78	6.75	6.73	8.22	9.03	8.64	8.39	8.06	7.61	7.00	6.91	6.84	6.84	8.69	7.58
	Cherrington	3.89	3.82	3.71	3.73	3.97	4.10	4.34	4.34	4.20	4.06	3.95	3.92	3.92	4.14	4.00
	Longdon	2.16	2.10	2.04	1.99	2.04	2.06	2.19	2.27	2.28	2.24	2.21	2.20	2.19	2.10	2.15

In order to demonstrate simulated groundwater level inter-model variability for the two extreme climate change scenarios 50L and 80H (shown previously as the extreme emission scenarios and from different time slices), regimes for groundwater levels at three boreholes (mean monthly levels over 2000-2003 period) are displayed in Figure 8.9. As with flows, the y axes scales are fixed (in this case to 10m) to enable comparisons between scenarios and sites.

In comparison to Figure 8.8 for river flows, the variability between models for the different scenarios is not so apparent. This may be due to the large y axes scales that were required due to the uncertainty in groundwater level simulations for the baseline period, before even assessing uncertainty between climate change scenarios.

Although the mean ranges in groundwater level variation between the six models shown in Table 8.5 indicate a progressive trend for the emission scenarios (low to high) and for each time slice (2050s and 2080s), the trend is not consistent at all boreholes as it was with river flow. For example, the range in intra-model actually decreases from the baseline period for Cherrington and Longdon (4.10m and 2.29m respectively) to 4.00m and 2.12m respectively, for the 80H scenario. In contrast, an increase in the intra-model uncertainty is seen at Gnosall, where the mean annual range increases from 6.62m to 7.58m (Table 8.5). These results highlight an important finding, that the variation in groundwater levels at different sites within the catchment differ more than for the signal apparent between climate change scenarios, as addressed in the subsequent Section 8.3.3.

With regard to any changes in seasonality or shape of the regimes only at Gnosall for the multi-criteria automatic calibration of the distributed model is it apparent there is an increase in the winter recharge for the 80H scenario when compared to the baseline period.

8.3.3. Uncertainty in climate change impacts from different scenarios in relation to uncertainty from different model setups

In order to put the results already discussed in this chapter into context, this section compares the uncertainty between models with the range of impacts of different climate change scenarios. Discussion of the climate data in Section 8.2, and presentation of the simulated results in Tables 8.4 and 8.5 have identified the 50L and 80H scenarios as the extreme climate change scenarios assessed in this work. Table 8.6 (river flows) and 8.7 (groundwater levels) provide a comparative summary of the results of these two scenarios and the baseline period. Each table provides the simulated mean daily flows or groundwater levels of the six calibrated models for the baseline and two climate change scenarios. The range (maximum mean monthly – minimum mean monthly) for the six models are also shown as a measure of model uncertainty. The percentage variation in simulated mean daily flow or groundwater level as a proportion of the mean baseline flow of groundwater level are also presented for the different scenarios. This gives an overall impression of the uncertainty between the six models, and can be compared to the percentage mean change in flow or groundwater level from the baseline period for the two scenarios. It is indicative of the overall impact of climate change for each

scenario. In both tables, the calculations have been undertaken at the four flow gauging stations and three groundwater level boreholes previously assessed.

Table 8.6 suggests an overall decrease in mean daily flow at all gauging stations. Aside from small variations, simulated changes are shown to be of a similar magnitude at all the gauging stations. The mean annual decreases simulated for the 50L scenario range between -13.5% to -17.9% increasing to -30.2% to 38.2% (approximately double that of 50L) for the 80H scenario. As these calculations are relative to the baseline calibrations, and constructed from a mean of the six models, this ensemble approach gives a good general indication of the impact of climate change as the intra-model uncertainty has been averaged.

The intra-model uncertainty at simulating the impacts of climate change on river flows for the different scenarios is shown in Table 8.6 by the range and percentage variation for the six hydrological models. The uncertainty range between the six models is shown to increase under future climate change scenarios, for all gauging stations. At the basin outlet, Walcot, the range of 10.4% between the six models increases to 16.2% for the 50L and 40.1% for the 80H scenarios, an overall increase from baseline to the 80H scenario of 29.7%. The increase in inter-model uncertainty with climate change scenarios is shown at Bailey Brook to be substantially greater. The baseline uncertainty (previously acknowledged in Chapter 7 as large) is 43.7%, this increases to 95.2% for the 80H scenario, an increase in uncertainty of 51.5%.

These results therefore suggest that at the basin outlet the uncertainty between models is of a similar magnitude to the impact of climate change signal, however, at internal gauging stations, the uncertainty between models is much greater than the climate change signal. These results are similar to the results shown by Wilby (2005) regarding parameter equifinality in a conceptual model of the Thames at Kingston. Here it was noted that during winter months, uncertainty from different parameter sets was of comparable magnitude to the uncertainty arising from the emission scenario. The work in this thesis furthers assessment of how uncertainty has been addressed as it highlights the intra model uncertainty at different sites internally within the catchment, and uses a primarily physically based, distributed model code rather than a conceptual model.

Similarly, large uncertainty between six different hydrological models assessed by Jiang et al (2007) for climate change impacts suggests that the results in this thesis are not unique, and that parameter uncertainty and model structure uncertainty require routine assessment before being used to assess the impacts of climate change and make water resource planning decisions.

Table 8.6. Compilation of the mean annual impact of climate change versus uncertainty between model simulations for four river flow gauging stations for the baseline and extreme climate change scenarios

	Walcot	Roden	Meese	Bailey Brook
Simulated MDF (m^3s^{-1})(mean of all six models)				
Baseline 2000-2003	6.90	1.84	1.17	0.26
50 Low scenario	5.96	1.52	1.01	0.22
80 High scenario	4.82	1.14	0.81	0.17
Range (m^3s^{-1}) (max-min) for all six models				
Baseline 2000-2003	0.72	0.17	0.14	0.12
50 Low scenario	0.97	0.30	0.22	0.13
80 High scenario	1.93	0.57	0.35	0.16
% variation in simulated MDF as a proportion of baseline MDF				
Baseline 2000-2003	10.4	9.3	11.6	43.7
50 Low scenario	16.2	20.0	21.6	59.6
80 High scenario	40.1	49.7	42.9	95.2
% mean change from baseline calibration (impact of climate change)				
50 Low scenario	-13.5	-17.2	-13.9	-17.9
80 High scenario	-30.2	-38.2	-30.8	-37.3

Table 8.7. Compilation of the mean annual impact of climate change versus uncertainty between model simulations for three groundwater level boreholes for the baseline and extreme climate change scenarios

	Gnosall	Cherrington	Longdon
Simulated MGWL (m) (mean of all six models)			
Baseline 2000-2003	82.31	57.05	53.07
50 Low scenario	81.95	56.79	52.93
80 High scenario	81.43	56.40	52.71
Range (m) (max-min) for all six models			
Baseline 2000-2003	6.56	4.10	2.29
50 Low scenario	6.74	4.06	2.25
80 High scenario	7.32	4.00	2.15
% variation in simulated MGWL as a proportion of baseline MGWL			
Baseline 2000-2003	7.97	7.19	4.32
50 Low scenario	8.22	7.15	4.24
80 High scenario	8.98	7.09	4.07
% mean change from baseline calibration (impact of climate change)			
50 Low scenario	-0.43	-0.45	-0.26
80 High scenario	-1.06	-1.14	-0.68

Table 8.7 suggests that the mean simulated groundwater levels from the six models indicate decreasing levels with climate change. This is expected and in line with other research in the UK that suggests a decrease in soil moisture (a related parameter) and hence groundwater recharge as an impact of climate change (Hulme et al., 2002). As with river flows, by using the mean value of the six models, the ensemble approach gives a fair and balanced assessment of the impact of climate change. At Gnosall, an overall mean decrease in levels of 0.88m is simulated for the 80H scenario, the largest decrease simulated for the three sites. The decline in groundwater levels is smaller at Longdon, 0.36m for the 80H scenario, and highlights the need of assessing a number of internal sites in distributed modelling, as internal variability (as with flows) has been shown.

In contrast to the simulated impacts of climate change, the uncertainty and range between models is not shown to consistently increase for the more extreme climate change scenarios such as 80H, as was the case for river flows. On the contrary, for Cherrington and Longdon, the mean range between the six models decreases for the 80H scenario when compared to the baseline period. Despite this, a progressive trend for each emission scenario that is magnified by the 80H time slice is shown.

The impact of climate change of groundwater levels is important as the main water supply in many parts of the world derives from groundwater. This is notably the case in the Tern catchment with the development of the Shropshire Groundwater Scheme (Section 3.8.2), a project being implemented that delivers water held in the sandstone aquifers to augment water levels further downstream in the River Severn at times of low flow.

The large uncertainties in simulated groundwater levels between the models for the baseline period calls into question the validity of using the models to simulate the impacts of climate change on groundwater levels (in this research). One of the main reasons for the discrepancy in projections (groundwater or for other variables) may be that simplistic assumptions are often made to represent the physical processes associated with hydrological systems (Goderniaux et al., 2009). This is particularly the case for models that simulate groundwater due to the complex nature and interactions within the sub-surface. The characteristic complex geology of the Tern catchment

highlights this further, and it is suggested that perhaps the 1km × 1km grid scale at which models have been developed in this thesis is too coarse for the catchment in question.

As (Goderniaux et al., 2009) suggest however, detailed physically based and spatially distributed simultaneous solving equation models (such as the distributed model in this research) that consider hydro-geological processes are likely to provide more realistic simulations of groundwater level fluxes. In theory this is due to a better representation of the whole system, as water flow in one domain is inter-connected with flow in other domains (Goderniaux et al., 2009).

It is acknowledged that the mean annual scale used to compare the impacts of climate change and uncertainty between models is relatively crude, but it does give an all round overview which has been the emphasis of this chapter. Additionally, this research does not account for the acknowledged uncertainties that have been shown in other studies Heuvelmans et al., (2004) for example, suggest that parameter values being transferred to simulate the effects of change may no longer be applicable for the new forcing data.

To summarise, the results of this study suggest that for river flows, uncertainty between simulations from different models increases when using the models to simulate the impacts of climate change, to equal or greater extents (for internal gauging stations) than the actual variability simulated by the climate change signal. Groundwater level simulations show a less clear pattern, where the uncertainty between the six models for the baseline period are so large that projected climate changes, even for the 80H scenario are minimal in comparison.

8.4. Summary

This chapter has furthered the uncertainty between models that were derived in Chapter 7 by using an example of a typical application of hydrological models – the assessment of the impacts of climate change on water resources. Rather than using one calibrated hydrological model to assess the impacts of climate change, all six hydrological models of differing spatial resolutions and calibration methods were forced with new input climate data that derived from the UKCIP02 data set. Eight climate change scenarios

were simulated for each model, with the 50 Low and 80 High selected as examples of the most extreme scenarios.

The objectives of this chapter were to assess the projected impacts of climate change. It was shown that both river flows and groundwater levels are projected to decrease (a drying trend) in the catchment. These results were noted to be similar findings of other climate change impact studies in the UK.

The second objective was to assess the uncertainty and range simulated between the six calibrated hydrological models for different climate change scenarios. For river flows it was shown that there was a consistent increase in uncertainty with climate change scenarios. The results further suggested that uncertainty between the models was greatest at internal gauging stations such as Bailey Brook. The intra-model uncertainty in groundwater level simulation for the baseline period was shown to be too large to sensibly quantify any projected increases in uncertainty with climate change.

Lastly the magnitude of intra-model uncertainty was compared to the range of projected impacts of climate change. It was shown that the uncertainty between models for river flow is as large as the signal derived from climate change scenarios. Widely variable groundwater level results for the baseline period led to suggestions that the results for groundwater levels could not be quantified as reliably. This work therefore highlights that water resource planners need to quantify the inter hydrological model uncertainties more rigorously before using simulated results of climate change to make decisions of how to mitigate the impacts of climate change.

Chapter 9

Conclusions and Recommendations

9.1. Summary of the research

By constructing a range of MIKE SHE hydrological catchment models for the 876.36 km^2 lowland Tern catchment in Shropshire, the research has addressed a range of key issues in hydrological modelling. These include uncertainties in model structure and differences in spatial representation, issues of model calibration and parameter equifinality, as well as confirming the need to assess distributed catchment models at a range of internal locations, for a range of performance statistics. In exploring the research questions outlined in Chapter 1, the main aim of this thesis was to use an ensemble modelling approach to assess whether a best modelling protocol could be developed which results in a model framework that could justifiably be considered better than the rest.

The structure of this thesis and division of chapters has been led by the principle steps undertaken in any hydrological modelling study. The thesis can be divided into four main parts; Chapters 1-3 that provided the background to the research, Chapters 4-6 that described model development and calibration, Chapter 7 that provided an assessment and evaluation of the models and Chapter 8 that used the models in a typical application to assess intra-model uncertainty when simulating the impacts of climate change.

9.2. Assessment of thesis aims and objectives

In Chapter 7, comparisons of the six optimally calibrated models from each protocol (spatial representation and calibration method) were presented. The chapter referred back to the first five research questions of the thesis (Chapter 1) and, in turn the methods used to explore each question were assessed. This was followed by a discussion of the results and evaluation of each of the research questions. Referring back to the research questions in Chapter 1, the conclusion that no modelling protocol could be justified as better than the rest was made for the following reasons:

1. What effects do different spatial representations have on model outputs?

With regard to model spatial complexity the results did not indicate (qualitatively or quantitatively) that either of the model representations performed better, especially at the basin outlet where all the modelling protocols resulted in similar simulation of river flow at Walcot. However, it was noted that in certain tributaries of the Tern, e.g. the Roden that is predominantly underlain by Mercia Mudstones and lower permeability geology than the characteristic Sherwood Sandstones of the catchment, the distributed model resulted in better performance. This result is likely due to the better non-homogenous representation of the different geology within this sub-basin. On the contrary, the Meese tributary performed well with the homogenous model set-up as the value for horizontal hydraulic conductivity (that was shown as a sensitive model parameter) better represented the sandstone geology.

It was suggested that the grid size 1km × 1km may have been too large to represent parameter values realistically. However, the trade off in grid size was principally due to a lack of available data at any higher resolution for the associated parameter values. It is noted in Section 7.4 that further research could involve the re-simulation of the models at a higher spatial resolution (to better represent the topography data, for example) and then assess the differences compared to the 1km x 1km grid cell used for the rest of the work in this thesis.

2. How do different methods of model calibration affect the performance of river flows and groundwater levels?

Performance of the different calibration methods indicated that it was possible to produce statistically acceptable models of the catchment using both manual and automatic methods. It was suggested that the multi-objective automatic calibration was the most robust as it included a range of measures (river flow and groundwater levels) at various sites in the optimisation process. The statistics for the model validation sites also demonstrated that the more comprehensive automatic calibration method resulted in a more robust calibration, as the groundwater level performance was better for the validation sites than the calibration sites. In this calibration method, the models had a higher mean score for sites that were not used in the calibration process.

The manual calibration method, especially for the homogenous model was also shown to work well for both river flows and groundwater level simulations. It was noted, however, that the calibration method used should reflect the aims of the research and consequently the spatial representation used in the model – with homogenous models seeking to assess river flow at the catchment outlet suited to manual calibration or automatic calibration optimising river flow at the basin outlet alone, whereas more complex distributed studies require a more rigorous calibration method.

3. How different are river flows and groundwater levels as a result of parameter equifinality?

As noted in the previous section, equifinality within the parameter space was assessed within Chapters 5 and 6 for the automatically calibrated models. Parameter equifinality was assessed by the selection of a sample of ten statistically similar models for each of the automatically calibrated models. It was demonstrated that for the models automatically calibrated at the basin outlet, the ten sampled models showed little differences in the output plots of simulations. Expectedly, the multi-criteria automatic calibration resulted in wider simulation bands, with the ten statistically very similar models having wider variation in the output plots. This was expected as the multi-criteria automatic calibration optimised on two objective functions for both river flow and groundwater level. The results of this part of the work did however demonstrate the uncertainty in model parameterisation; that different parameter sets can yield very statistically similar models.

4. To what extent do different performance statistics suggest different abilities of models?

The inclusion of a variety of statistics of model performance has also shown that performance statistics can be misleading, especially if used within the calibration process, such as the use of RMSE within the automatic calibration optimisations. It was therefore recommended that a range of performance statistics should be used to assess model performance. In this thesis, 'expert elicitation' in the form of qualitative model assessment and development of a further measure – the summary score, was used as a means to quantify the predominantly qualitative aspects of model performance such as ability of the models to simulate base flows, peak flows and the timing of hydrological events when compared to observed data.

5. How do measures of performance internally within the catchment, as well river flow at the basin outlet suggest different abilities of models?

In assessing model performance at a range of internal locations within the catchment it was shown that intra-model uncertainty ranged considerably between the different river flow gauging stations and groundwater level boreholes. The inclusion of the validation sites in the thesis demonstrated that the models also had ability (albeit slightly reduced compared to the calibration sites) in simulating river flow and groundwater levels at sites which were not calibrated.

In assessing performance at internal sites, it was also possible to highlight scaling effects within the models, for example in the smaller sub-basins of Bailey Brook and Potford Brook. In these sub-basins, intra-model uncertainty was a lot larger than in sub-basins that covered a larger area. The explanation for this is that over a smaller area, the model is more sensitive to the individual parameter values that have been set in the calibration. Likewise, over larger areas there is likely to be more heterogeneity that is 'averaged' out.

The majority of work compiling intra and inter-model uncertainty has previously been undertaken comparing models only at the basin outlet. This research is therefore considered to be some of the first in comparing internal model performance between different models developed using the same modelling code for the same catchment.

In Chapter 8, the uncertainty in climate change impacts between the same six best calibrated models from different protocols (assessed in Chapter 7) was demonstrated by forcing each hydrological model with perturbed meteorological data from eight UKCIP02 climate change scenarios. The purpose of this part of the research was to assess the implications of the intra-model uncertainty derived in Chapter 7, when used in a common application of hydrological models. Chapter 1 outlined that this part of the thesis addressed three further research questions:

6. What are the projected impacts of climate change simulated for the Tern catchment, using UKCIP02 data, for both river flows and groundwater levels at a range of locations?

The mean daily flow for the six models suggest for both the 2050s time slice with the low emission scenario and the 2080s time slice with the high emission scenario (the two extreme scenarios), that the catchment will become increasingly drier (-13.5% and -30.2%, respectively) compared to the mean baseline calibration of the six models. This drying trend was shown to be in line with other UK based projections of climate change (e.g. Arnell et al., 2003).

It was suggested that groundwater simulation for the calibration period would need to be improved before using the models to quantitatively assess the projected impacts of climate change. However, general decreases in groundwater levels were shown that varied in magnitude according to the models from the different modelling protocols.

7. What is the uncertainty and range of simulated river flows and groundwater levels between the six calibrated hydrological models for different climate change scenarios and two time-slices (2050s and 2080s)?

Intra-model uncertainty was also assessed for the six models for eight UKCIP02 climate change scenarios. The range in mean daily flow at the basin outlet for the six models for the calibration period was 10.4% of the mean of the six models. This increased to 40.1% for the 2080s High scenario. The increasing intra-model uncertainty between modelling protocols shown in this thesis indicate similar findings to other research that

has assessed inter-model uncertainty between different modelling codes (e.g. Jiang et al., 2007 and Wilby, 2005).

In addition, this research has further addressed the uncertainty internally within the catchment. It was shown that for the smaller, upstream tributary of Bailey Brook which already had a large intra-model uncertainty for the calibration period with a range of mean daily flows of 43.7% for the six models, that this increases to 95.2% for the 80H scenario. These results indicate that uncertainty assessed at the basin outlet cannot necessarily be considered to represent the catchment internally, an important consideration for water managers.

With regard to groundwater levels, the intra-model uncertainty derived for the different hydrological modelling protocols for the calibration period (2000-2003) was shown to be large. When assessing any changes to the intra-model uncertainty with projected climate changes the results were comparatively negligible, although the largest intra-model uncertainty was shown at Gnosall – a borehole described throughout the thesis as the most difficult to simulate groundwater levels at, due to its proximity to the catchment boundary.

8. What is the magnitude of the intra-model uncertainty compared to the projected impacts of climate change, and how important is intra-model variability when using the models to simulate the potential effects of climate change?

As described in the summaries for research questions six and seven, the results of this thesis suggest that intra-model uncertainty is of similar magnitude to the potential impacts of climate change. This result is important when considering that projections of the impacts of climate change derived from hydrological modelling are often used to inform water managers in the decision making process. Aside of increasing intra-model uncertainty between the most extreme scenarios assessed (50L and 80H), it is important to note that the different modelling protocols suggested different projected changes to the annual river flow regimes, for example in the Roden tributary. The manually calibrated homogenous model did not suggest any changes to the regime (but general decreases in the volume of water), whereas the automatic calibrations of the distributed model suggested that the month of peak river flow changes from November (baseline calibration) to February (80H scenario). This thesis has therefore demonstrated the

uncertainty that would arise from basing water management decisions on only one of these hydrological models, and highlights the need for rigorous uncertainty analyses of different hydrological modelling protocols.

9.3. Contributions to knowledge and recommendations

In addition to the conclusions that have been given for each of the thesis research questions, the contributions to knowledge as a result of this research can also be drawn from within the thesis:

1) Chapter 3 provided a detailed review of the Tern catchment. The chapter presented the secondary data that was used in the research and included primary analyses of these data. At the end of the chapter, the available data were listed within one table for easy reference for those wishing to know where data for the catchment can be found. A conceptual figure for the catchment was also constructed that provides an overview of the differences in spatial representation for key processes that influence the hydro-geology.

> **Recommendation:** The catchment review and conceptual figure may be useful for anyone else undertaking research in the catchment. It is also recommended that others undertaking catchment modelling should provide a similar conceptual model as it serves as an easy reference for the modeller and other readers to understand the key features of the catchment.

2) Detailed sensitivity analyses have been undertaken in Chapters 5 and 6 that not only enabled the most sensitive model parameters to be selected for further calibration, but also to check and confirm the internal operation of the model and that it is operating as should be expected. This thesis has documented a method of how this can be undertaken using both manual (Chapter 5) and automatic methods (Chapter 6). In the case of the distributed model, the sensitivity analysis provided an uncertainty analysis of the model response to parameter variation – with individual profiles for each of the 22 assessment river flow gauges or groundwater level boreholes. The presentation of these individual profiles demonstrated the differences in model response at the different sites.

Recommendation: It is suggested that the sensitivity analysis is a key phase of the modelling protocol and should not be overlooked, especially as this stage enables the modeller to learn a great deal about the internal operation of the model (especially important in distributed modelling).

3) The results of this work demonstrated that if a particular statistic has been used within the calibration as part of the optimisation process (e.g. the RMSE in the automatic calibrations), then ideally this performance indicator should not solely be used to assess model performance. For a more robust assessment it is better to use alternative indicator. As such, an independent performance measure that does not depend on being included within the calibration optimisation has been developed: the summary score. This measure provides an independent assessment of model performance based on the predominantly qualitative information that can be drawn from simulation result plots.

Recommendation: It is suggested that a range of performance statistics should be used to assess model performance, at that those that are optimised during calibration are unfairly improved compared to other statistics. As a result, it is important to assess model performance both quantitatively and qualitatively by looking at the result plots of simulations. It is suggested a methodology similar to that used in this research to calculate the summary score is a useful way of independently assessing performance.

4) Despite having developed and tested a range of modelling protocols based on different spatial representation (a homogenous approach and a more distributed approach based on the available data) each with three calibration methods, many of the issues that have been described with each model in Chapters 5 and 6 are the same in each protocol despite the different set-ups. For example, groundwater levels at Gnosall near the catchment boundary were always poorly simulated in all models. A contribution of this research is therefore in demonstrating that regardless of the effort made to set up a distributed model or time taken in calibration, it does not overcome many of the larger issues (e.g.adequate boundary representation).

Recommendation: When deciding on the best modelling protocol for the research in question, it is important to minimise the larger issues that may influence overall modelling results (that will not be influenced by the choice of spatial representation or calibration method, for example). These issues may include boundary information for sub-surface water movement within the aquifer, abstraction or discharge data, or representing the river network within the catchment to the best available level. If these issues cannot be overcome, then it is necessary to acknowledge them as best possible.

5) For the calibration sites used in the model, it was not possible to choose a best model/protocol. It was shown that statistically, very similar performing models can be developed with different model set-ups and calibration methods – especially at the basin outlet, Walcot. The research therefore demonstrates statistical equifinality within the modelling protocol.

Recommendation: The results of this research suggested that despite similar performance statistics, intra-model uncertainty (in simulated plots) was larger at internal sites within the catchment. Additionally, the intra-model uncertainty increased when the models were applied to assess the impacts of climate change. It is therefore recommended that intra-model uncertainty should be further addressed by others. At present, a wide body of literature exists that assesses the uncertainty of parameterisation, and different types of data input (e.g. different methods of rainfall data). However, the uncertainties attributed to the choices made by the modeller during the development of the modelling protocol phase have been more widely overlooked. It is recommended that this uncertainty should be specifically addressed when using the models for assessment impacts (e.g. climate change or land use change) as this research demonstrates it is during this stage that the intra-model uncertainty is greatest.

The key contribution of this research that is documented throughout the thesis is that the research includes comparisons of different calibration methods and spatial representations of data in hydrological modelling. Although specific conclusions are drawn on these issues in the previous sections, this issue more widely is a contribution to knowledge since very few studies assess this intra-model uncertainty with different

models of the same catchment, or using different set ups of the same modelling code, in this research the modelling code being MIKE SHE.

9.4. Recommendations for further research

Further research that could refine the hydrological models in this thesis could include a further exploration of the evapotranspiration representation in the model. It would be interesting to drive the model with alternative input evapotranspiration data to assess its influence on the sensitivity of the calibrated models. For example, Kingston et al., (2009) assessed different methods of input PET at the global scale and suggested assessment at more local catchment scales would be appropriate.

Additionally, refining the hydrological models with aquifer boundary information in the saturated zone could also be a further area of research specific to the models developed. For example, boundary conditions from regional groundwater models (e.g. from the British Geological Survey). Licensed abstraction and effluent return data within the catchment that exist but were not available for this research may also improve the water balance calculations in the models, and therefore improve resulting hydrological simulations.

This thesis has primarily assessed model uncertainty as a result of the different choices that can be made during the modelling protocol. Further assessment of uncertainty could be undertaken by ascertaining how potentially erroneous or suspect input data influence model results and performance statistics. Although not assessed within the specific research questions of this thesis, this research highlighted potentially suspect rainfall data that could not be confirmed as erroneous, and so was not adjusted. The use of these data were justified and used in the simulations but could potentially have influenced the overall model performance statistics such as Nash-Sutcliffe NSE. In the case of a very high daily rainfall value that is suspect, further research could assess the impact this has within the model; on nearby river flow, river flow at the catchment outlet as well as other measures such as groundwater levels.

This research has used the modelling code MIKE SHE. Although this code has been shown to be successful and suitable for the research purposes, it could also be

worthwhile to compare results from other modelling codes. One of the benefits of this thesis is that all of the models were developed with the same code, therefore reducing inter-model uncertainty. Further work may seek to compare different modelling protocols within the same framework – such as with the semi-distributed SWAT (Arnold et al., 1993, 1998) or SLURP (Kite, 1995) models, or distributed WATFLOOD (Kouwen et al., 1993) model for example.

It is acknowledged that only two scenarios of spatial representation were developed in this research. An instinctive suggestion may be to improve the distributed model with a finer grid although this may be logistically prohibited due to the amount of data that it would require. Other options for research to assess uncertainty from modelling protocol may therefore be to investigate different distributions of input data. For example, assessing uncertainty between simulations that distribute geological spatial data, or soil data, into different numbers of classes. The distributed model represented the saturated zone with five classes. The method adopted resulted in the necessary specification of 20 attribute parameters for the saturated zone. It was described that the method resulted in some of the spatial variation within the soil types being included; however, it is acknowledged that the real-world heterogeneity in soil types means that the spatial distribution was relatively crude. It would be possible to further assess different representations of the unsaturated zone within the model to assess the impact on performance, for example, by including a larger number of classes.

This research has been successful in bringing together a wide body of data that has been collected by different organisations as a result of different projects such as LOCAR. As distributed models require large amounts of data for a range of variables and parameters it would not have been possible to undertake the amount of hydrological modelling in this thesis as well as primary data collection and catchment wide monitoring within the scope of this PhD research. Having developed a range of catchment models and a MIKE 11 river model, this work highlights that it is possible to synthesise existing, archived data from difference sources and produce new and useful outputs with it.

It is also acknowledged that the Tern catchment has been more widely monitored in the past than other catchments in the UK and overseas. Although this was a governing

factor in selecting the catchment, reproducibility of the method used in this thesis may be difficult in many other catchments without additional data collection.

It is also notable that in many cases in this research, it was not that spatial data resolution was insufficient for a distributed model of finer grid sizes, but that attribute values were lacking for many of the parameters associated with the data that had already been collected, such as site specific saturated zone parameters. Further work may therefore include the collection of specific parameter values, although this is characteristically time consuming and expensive to routinely prohibit collection at the catchment scale. It is suggested that further research comparing intra-model uncertainty could usefully be undertaken on a smaller catchment. It could be worthwhile to compare results within a sub-basin of the Tern catchment as well as comparing results from another catchment to assess whether the results found in this thesis are comparable.

As higher resolution topography data were available but re-sampled to a 1km x 1km lower resolution to best fit the resolution of the associated parameters (and in a trade off with the number of simulations required to undertake the work versus computer simulation time), it would also be possible to undertake simulations at a higher resolution but without increasing the spatial variability in the associated attributes for which there is no higher resolution available. For example, the hydrological models could be simulated on a finer 100m x 100m grid to assess the differences in simulation compared to that using the 1km x 1km grid that has been adopted.

The work in assessing the impacts and intra-model uncertainty of climate change could further be developed with the inclusion of a wider ensemble of calibrated models. For example, the suites of ten calibrated models that were derived in Chapters 5 and 6 as a result of automatic calibrations could each be forced with the same climate change data so as to more fully assess the uncertainty from different parameterisations of the same modelling protocols. This was not undertaken due to time restrictions and clarity in presenting the results. However, if it were undertaken then the ensembles would be comprised of a larger set of models in which to refine the uncertainty when simulating the impacts of climate change for the range of models (a total of 320 simulations – ten for each of the four automatic calibrations, for eight climate change scenarios).

The UKCIP02 data that was used in the research has now largely been superseded by the UKIP09 data described briefly in Chapter 9. It was initially intended that these data could be used in this thesis but due to the delay in their release it was not possible to do this. Further research could use these data to update the results in this thesis, and could additionally be used to compare the results between the two datasets.

References

Abbott, M. B., Bathurst, J. C., Cunge, J. A., Oconnell, P. E., and Rasmussen, J., 1986. An Introduction to the European Hydrological System - Systeme Hydrologique Europeen, She .1. History and Philosophy of A Physically-Based, Distributed Modeling System. Journal of Hydrology, 87 (1-2), 45-59.

Abbott, M. B., Bathurst, J. C., Cunge, J. A., Oconnell, P. E., and Rasmussen, J., 1986. An Introduction to the European Hydrological System - Systeme Hydrologique Europeen, She .2. Structure of A Physically-Based, Distributed Modeling System. Journal of Hydrology, 87 (1-2), 61-77.

Abrahamsen, P., and Hansen, S., 2000. Daisy: An Open Soil-Crop-Atmosphere System Model. in Environ. Model. Software v15, 313-330.

Abu El-Nasr, A., Arnold, J. G., Feyen, J., and Berlamont, J., 2005. Modelling the hydrology of a catchment using a distributed and a semi-distributed model. Hydrological Processes, 19 (3), 573-587.

Adams, B., Peach, D. and Bloomfield, J. P., 2003. The LOCAR Hydrogeological Infrastructure for the Tern Catchment. IR/03/180, British Geological Survey, Keyworth, Nottingham.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M., 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements, Irrigation and Drainage Paper 56, Food and Agriculture Organization of the United Nations, Rome.

Altenford, C. T., 1992. Using a neural network for soil moisture predictions. ASAE Microfiche No. 92-3557. St. Joseph, MI: ASAE.

Andersen, J., Sandholt, I., Jensen, K.H., Refsgaard, J.C., and Gupta, H., 2002. Perspectives in using a remotely sensed dryness index in distributed hydrological models at the river-basin scale, Hydrological processes,16, 2973-2987.

Andersen, J., Dybkjaer, G., Jensen, K. H., Refsgaard, J. C., and Rasmussen, K., 2002. Use of remotely sensed precipitation and leaf area index in a distributed hydrological model. Journal of Hydrology, 264 (1-4), 34-50.

Andersen, J., Refsgaard, J. C., and Jensen, K. H., 2001. Distributed hydrological modelling of the Senegal River Basin - model construction and validation. Journal of Hydrology, 247 (3-4), 200-214.

Apaydin, H., Anli, A. S., and Ozturk, A., 2006. The temporal transferability of calibrated parameters of a hydrological model. Ecological Modelling, 195 (3-4), 307-317.

Arheimer, B. and Brandt, M., 1998. Modelling nitrogen transport and retention in the catchments of southern Sweden. Ambio 27(6), 471-480.

Arnell, N. W. and Reynard N. S., 2000. Climate change and UK hydrology. In: M.C. Acreman, Editor, The Hydrology of the UK: A Study of Change, Routledge, London, 3-29.

Arnell, N. W., 1998. Climate Change and Water Resources in Britain, Climate Change, 39, 83-110.

Arnell, N. W., 1999a. The effect of climate change on hydrological regimes in Europe: a continental perspective. Global Environment Change 9, 5-23.

Arnell, N. W., 1999b. A simple water balance model for the simulation of streamflow over a large geographic domain. Journal of Hydrology, 217 (3-4), 314-335.

Arnell, N. W. and Reynard, N. S., 1996. The effects of climate change due to global warming on river flows in Great Britain. Journal of Hydrology, 183 (3-4), 397-424.

Arnell, N. W., 2003. Relative effects of multi-decadal climatic variability and changes in the mean and variability of climate due to global warming: future streamflows in Britain. Journal of Hydrology, 270 (3-4), 195-213.

Arnold, J. G., Allen, P. M. and Bernhardt, G., 1993. A comprehensive surface-groundwater flow model. Journal of Hydrology, 142 (1-4), 47-69.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. and Williams, J. R., 1998. Large area hydrologic modeling and assessment Part I: model development. Journal of the American Water Resources Association 34 (1).

Avery, B. W., 1980. Soil Classification for England and Wales (Higher Categories). Soil Survey Technical.

BADC, 2008. British Atmospheric Data Centre. Minimum and Maximum temperature data, accessed August 2008, http://badc.nerc.ac.uk/data/dataset_index/

Bathurst, J. C., 1986. Physically based distributed modeling of upland catchment using the Systeme Hydrologique Europeen. Journal of Hydrology, 87, 79-123.

Becker, A., 1992. Criteria for a hydrologically sound structuring of large scale land surface process models. In: O'Kane J.O.P. (Ed.). Advances in Theoretical Hydrology. Elsevier, Amsterdam, 97-112.

Becker, A., 1995. Problems and progress in macroscale hydrologic modelling. In: Feddes, R. A. (Ed.). Space and Time Scale Variability and Interdependencies in Hydrological Processes, Cambridge University Press, 135.

Becker, A. and Braun, P., 1999. Disaggregation, aggregation and spatial scaling in hydrological modelling. Journal of Hydrology, 217 (3-4), 239-252.

Bergström, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments. Ph.D. Thesis. SMHI Reports RHO No. 7, Norrköping.

Bergström, S., Harlin, J. & Lindström, G., 1992. Spillway design floods in Sweden. I: New guidelines. Hydrological Sciences Journal, 37 (5), 505-519.

Beven, K. J. and Kirkby M. J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Science Bulletin, 24 (1), 43-69.

Beven, K. J., 1996. A discussion of distributed hydrological modelling. In: Abbott, M. B., Refsgaard, J. C. (Eds.), Distributed Hydrological Modelling. Kluwer Academic, 255-278.

Beven, K. and Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes, 6, 279–298.

Beven, K. J., Calver, A. and Morris, E. M., 1987. The Institute of Hydrology Distributed Model, Institute of Hydrology, Report No. 98. Wallingford.

Beven, K., 1989. Changing Ideas in Hydrology - the Case of Physically-Based Models. Journal of Hydrology, 105 (1-2), 157-172.

Beven, K., 2001. On hypothesis testing in hydrology. Hydrological Processes, 15 (9), 1655-1657.

Beven, K., 2002. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. Hydrological Processes, 16 (2), 189-206.

Biftu, G. F. and Gan, T. Y., 2001. Semi-distributed, physically based, hydrologic modeling of the Paddle River Basin, Alberta, using remotely sensed data. Journal of Hydrology, 244 (3-4), 137-156.

Bingeman, A. K., Kouwen, N. and Soulis, E.D., 2006. Validation of the hydrological processes in a hydrological model. Journal of Hydrologic Engineering, 11 (5), 451-463.

Blackie, J. R. and Eles, C. W. O., 1985. Lumped Catchment Models, in M. G. Anderson and T. P. Burt (eds), Hydrological Forecasting, Wiley, New York, U.S.A.

Blöschl, G. and Sivapalan, M., 1995. Scale issues in hydrological modeling: a review. Hydrological Processes, 9, 251–290.

Boegh, E., Thorsen, M., Butts, M. B., Hansen, S., Christiansen, J. S., Abrahamsen, P., Hasager, C. B., Jensen, N. O., van der Keur, P., Refsgaard, J. C., Schelde, K., Soegaard, H., and Thomsen, A., 2004. Incorporating remote sensing data in physically based distributed agro-hydrological modelling. Journal of Hydrology, 287 (1-4), 279-299.

Boorman, D. B., Hollis, J. M. and Lilly, A., 1995. Hydrology of soil types: a hydrologically-based classification of the soils of the United Kingdom. Institute of Hydrology Report No. 126, Wallingford, Oxfordshire, 134.

Boughton, W., 2006. Calibrations of a daily rainfall-runoff model with poor quality data. Environmental Modelling & Software, 21 (8), 1114-1128.

Brath, A., Montanari, A. and Toth, E., 2004. Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model, Journal of Hydrology, 291, 232-253.

Brun, A. and Engesgaard, P., 2002a. Modelling of transport and biogeochemical processes in pollution plumes: literature review and model development, in Journal of Hydrology, 256, 211-227.

Brun, A., Engesgaard, P., Christensen, T. H. and Rosbjerg, D., 2002b. Modelling of transport and biogeochemical processes in pollution plumes: Vejen landfill, Denmark, in Journal of Hydrology, 256, 228-247.

Canadell, J., Jackson, R. B., Ehleringer, J. R., Mooney, H. A., Sala, O. E. and Schulze, E. D., 1996. Maximum rooting depth for vegetation types at the global scale. Oecologia, 108, 583-595.

Carpenter, T.M., Georgakakos, K.P., Sperfslage, J.A., 2001. On the parametric and NEXRAD-radar sensitivities of a distributed hydrologic model suitable for operational use. Journal of Hydrology, 254, 169–193.

Carpenter, T. M. and Georgakakos, K. P., 2004a. Continuous streamflow simulation with the HRCDHM distributed hydrologic model. Journal of Hydrology, 298 (1-4), 61-79.

Carpenter, T. M. and Georgakakos, K. P., 2004b. Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model. Journal of Hydrology, 298 (1-4), 202-221.

Caudill, M. and Butler, C., 1992a. Understanding Neural Networks: Volume 1: Basic Networks. Cambridge, Massachusetts: The MIT Press.

Caudill, M. and Butler, C., 1992b. Understanding Neural Networks: Volume 2: Advanced Networks. Cambridge, Massachusetts: The MIT Press.

Chaplot, V., Saleh, A., and Jaynes, D. B., 2005. Effect of the accuracy of spatial rainfall information on the modeling of water, sediment, and NO3-N loads at the watershed level. Journal of Hydrology, 312 (1-4), 223-234.

Chapman, L. and Thornes, J. E., 2003. The use of geographical information systems in climatology and meteorology. Progress in Physical Geography, 27 (3), 313-330.

Chaubey, I., Cotter, A. S., Costello, T. A., and Soerens, T. S., 2005. Effect of DEM data resolution on SWAT output uncertainty. Hydrological Processes, 19 (3), 621-628.

Christiaens, K. and Feyen, J., 2000. The influence of different methods to derive soil hydraulic properties on the uncertainty of various model outputs of a distributed hydrological model. Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere, 25 (7-8), 679-683.

Christiaens, K. and Feyen, J., 2001. Analysis of uncertainties associated with different methods to determine soil hydraulic properties and their propagation in the distributed hydrological MIKE SHE model. Journal of Hydrology, 246 (1-4), 63-81.

Christiaens, K. and Feyen, J., 2002a. Constraining soil hydraulic parameter and output uncertainty of the distributed hydrological MIKE SHE model using the GLUE framework. Hydrological Processes, 16 (2), 373-391.

Christiaens, K. and Feyen, J., 2002b. Use of sensitivity and uncertainty measures in distributed hydrological modeling with an application to the MIKE SHE model. Water Resources Research, 38 (9).

Christiansen, J, S., Thorsen, M., Clausen, T., Hansen, S. and Refsgaard, J, C., 2004. Modelling of macropore flow and transport processes at catchment scale, Journal of Hydrology. 299, 136-158.

Cosby, B. J., Hornberger, G. M., Clapp, R. B. and Ginn, T. R., 1984. A statistical exploration of the relationship of soil moisture characteristics to the physical properties of soils. Water Resources Research 20/6, 682-690.

Crawford, N. H. and Linsey, R.K., 1966. Digital Simulation in Hydrology: Stanford Watershed Model IV, Technical Report 39, Department of Civil Engineering, Stanford University, California.

Daniell, T. M., 1991. Neural networks - Applications in hydrology and water resources engineering. Proceedings, International Hydrology and Water Symposium, 3, 797-802. Nat. Conf. Publ. 91/22, Institute of Engineering, Australia, Barton, ACT, Australia.

DHI-WE, 2005. MIKE SHE Water Movement: User manual, Danish Hydraulic Institute – Water and Environment, Hørsholm, Denmark.

Dingman, S. L., 1994. Physical Hydrology. Prentice Hall, Englewood Cliffs, NJ, 575.

Diskin, M. H., 1970. On the computer evaluation of Thiessen weights. Journal of Hydrology, 11 (1), 69-78.

Duan, Q., Gupta, V. K. and Sorooshian, S., 1993. A shuffled complex evolution approach for effective and efficient global minimization. Journal of Optimization Theory and Applications., 76, 501-521.

Duncan, M. R., Austin, B., Fabry, F., and Austin, G. L., 1993. The Effect of Gauge Sampling Density on the Accuracy of Streamflow Prediction for Rural Catchments. Journal of Hydrology, 142 (1-4), 445-476.

Dunn, S. M. and Lilley, A., 2001. Investigating the relationship between a soils classification and the spatial parameters of a conceptual catchment scale hydrological model, Journal of Hydrology, 252, 157-173.

Dunne, T., 1983. Relation of field studies and modelling in the prediction of storm runoff. Journal of Hydrology, 65, 25-48.

Edina Digimap© Crown Copyright/database right 2007. An Ordnance Survey/EDINA supplied service.

Environment Agency, 2009, Shropshire Groundwater Scheme. Using groundwater to maintain flow in the River Severn, accessed January 2009. http://publications.environment-agency.gov.uk/pdf/GEMI0205BINJ-e-e.pdf

Ewen, J., O'Donnell, G., Burton, A. and O'Connel,l P. E., 2006. Errors and uncertainty in physically-based rainfall–runoff modelling of catchment change effects. Journal of Hydrology, 330 (3–4), 641-650.

Feyen, L., Vazquez, R., Christiaens, K., Sels, O. and Feyen, J., 2000. Application of a distributed physically-based hydrological model to a medium size catchment. Hydrology and Earth System Sciences, 4 (1), 47-63.

Flickin, D. L., Luo, Y. Z., Luedeling, E. and Zhang, M., 2009. Climate change sensitivity assessment of a highly agricultural watershed using SWAT, Journal of Hydrology, 374, 16–29.

Fleming, G., 1975. Computer Simulation Techniques in Hydrology. Elsevier, New York, USA.

Fowler, H. J., Kilsby, C. G. and Stunell, J., 2007. Modelling the impacts of projected future climate change on water resources in northwest England. Hydrology and Earth System Sciences 11(3), 1115–1126.

Fowler, H. J., Kilsby, C. G. and O'Connell, P. E., 2003. Modeling the impacts of climatic change and variability on the reliability, resilience and vulnerability of a water resource system, Water Resources Research. 39, 1222.

Franchini, M., Wendling, J., Obled, C., and Todini, E., 1996. Physical interpretation and sensitivity analysis of the TOPMODEL. Journal of Hydrology, 175 (1-4), 293-338.

Freer, J., Beven, K., and Ambroise, B., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resources Research, 32 (7), 2161-2173.

Freeze, R. A. and Cherry, J. A., 1979. Groundwater. Prentice-Hall, Inc. Englewood Cliffs, NJ, 604

French, M., Krajewski, W. F. and Cuykendall, R. R., 1992. Rainfall forecasting in space and time using a neural network. Journal of Hydrology, 137, 1-31.

Fuller, R. M., Smith, G. M., Sanderson, J. M., Hill, R.A. and Thomson, A.G., 2002. Land Cover Map 2000: construction of a parcel-based vector map from satellite images. The Cartographic Journal, 30 (1), 15–25.

Georgakakos, K. P., Sperfslage, J. A. and Guetter, A. K., 1996. Operational GIS-Based Models for NEXRAD Radar Data in the US Proceedings International

Conference on Water Resources and Environmental Research: Towards the 21st Century, October 29-31, Kyoto, Japan.

Goderniaux, P., Brouyère, S., Fowler, H. J., Blenkinsop, S., Therrien, R., Orban, P. and Dassargues. A., 2009. Large scale surface-subsurface hydrological model to assess climate change impacts on groundwater reserves. Journal of Hydrology, 373, 122-138.

Graham, D. N. and Butts, M. B., 2005. Flexible, integrated watershed modelling with MIKE SHE. In Watershed Models, Eds. Singh V. P. and Frevert D. K., 245-272.

Gupta, H. V., Sorooshian, S. and Yapo, P. O., 1998. Toward Improved Calibration of Hydrologic Models: Multiple and Noncommensurable Measures of Information. Water Resources Research, 34 (4), 751-763.

Gustard, A., Bullock, A. and Dixon, J. M., 1992. Low Flow Estimation in the UK. Report no. 108, Centre for Ecology and Hydrology (previously Institute of Hydrology), Wallingford, UK.

Hamby, D. M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. Environmental Monitoring and Assessment, 32, 135–154.

Hamon, W.R., 1961. Estimating potential evaporation, Division, Journal of Hydrology, (Ed.), Proceedings of the American Society of Civil Engineers, 107-120.

Hansen, S., Jensen, H. E., Nielsen, N. E., Svendsen, H., 1990. DAISY: Soil, Plant Atmosphere System Model. NPO Report No. A 10. The National Agency for Environmental Protection, Copenhagen, 272.

Hargreaves, G. H. and Samani Z. A., 1982. Estimating potential evapotranspiration, Journal of Irrigation and Drainage Engineering, ASCE, 108(IR3), 223-230.

Heneggeler, J. C., Samani, Z., Flynn, M. S. and Zeitler, J. W., 1996. Evaluation of various evapotranspiration equations for Texas and New Mexico. Proceedings of Irrigation Association International Conference, San Antonio, Texas.

Henriksen, H. J., Troldborg, L., Højberg, A. J. and Refsgaard, J. C., 2008. Assessment of exploitable groundwater resources of Denmark by use of ensemble resource indicators and a numerical groundwater–surface water model. Journal of Hydrology, 348, 224-240.

Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C., and Madsen, B., 2003. Methodology for construction, calibration and validation of a national hydrological model for Denmark. Journal of Hydrology, 280 (1-4), 52-71.

Heuvelmans, G., Muys, B. and Feyen, J., 2004. Evaluation of hydrological model parameter transferability for simulating the impact of land use on catchment hydrology. Physics and Chemistry of the Earth, Parts A/B/C 29 (11/12), 739-747.

Hobbs, P. R. N., Hallam, J. R., Forster, A., Entwisle, D. C., Jones, L. D., Cripps, A. C., Northmore, K. J., Self, S. J. and Meakin, J. L., 2002. Engineering geology of

British rocks and soils: mudstones of the Mercia Mudstone Group. British Geological Survey Research Report, RR/01/02, 106.

Holland, J. H., 1975. Adaptation in natural and artificial systems. Ann Arbor, MI: University of Michigan Press.

Horritt, M. S. and Bates, P. D., 2001. Effects of spatial resolution on a raster based model of flood flow. Journal of Hydrology, 253 (1-4), 239-249.

Howard, A. S., Warrington, G., Ambrose, K. and Rees, J. G., 2008. A formational framework for the Mercia Mudstone Group (Triassic) of England and Wales. British Geological Survey Research Report, RR/08/04.

Hough, M., 2003. An Historic Comparison between the Met Office Surface Exchange Scheme-Probability Distributed Model (MOSES-PDM) and the Met Office Rainfall and Evaporation Calculation System (MORECS). Meteorological Office Report.

Hulme, M., Jenkins, G. J., Lu, X., Turnpenny, J. R., Mitchell, T. D., Jones, R. G., Lowe, J., Murphy, J. M., Hassell, D., Boorman, P., McDonald, R. and Hill, S., 2002. Climate Change Scenarios for the United Kingdom: The UKCIP02 Scientific Report, Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich, UK, 120.

Hutchinson, P., 1970. A contribution to the problem of spacing raingauges in rugged terrain, Journal of Hydrology, 12 (1), 1-14.

HYACINTS, 2009. Hydrological Modelling for Assessing Climate Change Impacts at different Scales, accessed January 2009, http://hyacints.dk/main_uk/main.html

IPCC., 2001. Climate change 2001: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change. In: McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J. and White, K. S. (eds.) Cambridge University Press, Cambridge, 1032.

Jain, S. K., Storm, B., Bathurst, J. C., Refsgaard, J .C., and Singh, R. D., 1992. Application of the SHE to Catchments in India - Part 2: Field Experiments and Simulation Studies with the SHE on the Kolar Subcatchment of the Narmada River, Journal of Hydrology, 140, 25-47.

Jayatilaka, C. J., Storm, B. and Mudgway, L. B., 1998. Simulation of water flow on irrigation bay scale with MIKE SHE, Journal of Hydrology, 208, 108-130.

Jenkins, G. J., Perry, M. and Prior, J., 2009. The climate of the UK and recent trends. Revised edition, Met Office Hadley Centre, Exeter.

Jensen, M. E., Burman, R. D. and Allen, R. G. (ed)., 1985. Evapotranspiration and Irrigation Water Requirements. ASCE Manuals and Reports on Engineering Practices No. 70., American Society of Civil Engineers, New York.

Jiang, T., Chen, Y. D., Xu, C-Y., Chen, X., Chen, X., 2007. Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. Journal of Hydrology, 336, 316-333.

Johnson, A. I., 1967. Specific yield - compilation of specific yields for various materials. U.S. Geological Survey Water Supply Paper 1662-D, 74.

Jutman, T., 1992. Production of a new runoff map of Sweden. Proceedings of Nordic Hydrological Conference, 4-6 August, Alta, Norway, NHP report, nr. 30, 643-651.

Kabat, P., Hutjes, R. A. W. and Feddes, R. A., 1997. The scaling characteristics of soil parameters: From plot scale heterogeneity to subgrid parameterization. Journal of Hydrology, 190, 363-396.

Kang M. S., Park S. W., Lee J. J. and Yoo K. H., 2006. Applying SWAT for TMDL programs to a small watershed containing rice paddy fields. Agricultural Water Management, 79 (1), 72-92.

Karvonen, T., Koivusalo, H., Jauhiainen, M., Palko, J., and Weppling, K., 1999. A hydrological model for predicting runoff from different land use areas. Journal of Hydrology, 217 (3-4), 253-265.

Kingston, D. G. and Taylor, R. G., 2010. Projected impacts of climate change on groundwater and stormflow in a humid, tropical catchment in the Ugandan Upper Nile Basin, Hydrology Earth System Science Discussions, 7, 1913-1944.

Kingston, D.G., Todd, M.C., Taylor, R.G., Thompson, J.R. and Arnell, N.W. 2009. Uncertainty in the estimation of potential evapotranspiration under climate change. Geophysical Research Letters 36, L20403.

Kirkby, M. J., Naden, P. S., Burt, T. P. and Butcher, D. P., 1993. Computer simulation in physical geography (2nd edition), Chichester: Wiley.

Kite, G. W., 1995. The SLURP model. In Singh, V. P. (Ed.). Computer Models of Watershed Hydrology. Water Resources Publications, Colorado, 521-562 (Chapter 15).

Kite, G. W. and Droogers, P., 1999. Irrigation modelling in the context of basin water resources. Journal Water Resources Development, 15, 1/2, 43-54.

Kite, G. W., 1997. Simulating Columbia River flows with data from regional-scale climate. Water Resources Research, 33 (6), 1275-1285.

Kite, G. W., Dalton, A. and Dion, K., 1994. Simulation of streamflow in a macro-scale watershed using GCM data. Water Resources Research, 30 (5), 1546-1559.

Klemes, V., 1983. Conceptualization and scale in hydrology. Journal of Hydrology, 65, 1-23.

Klemes, V., 1986. Operational Testing of Hydrological Simulation-Models. Hydrological Sciences Journal-Journal des Sciences Hydrologiques, 31 (1), 13-24.

KNMI, 2008. Climate Explorer, accessed June 2008, http://climexp.knmi.nl/

Koren, V. I., Finnerty, B. D., Schaake, J. C., Smith, M. B., Seo, D-J. and Duan, Q-Y., 1999. Scale dependencies of hydrologic models to spatial variability of precipitation. Journal of Hydrology, 217 (3-4), 285-302.

Kouwen, N., Soulis, E.D., Pietroniro, A., Donald, J. and Harrington R.A., 1993. Grouping Response Units for Distributed Hydrologic Modelling. ASCE Journal of Water Resources Management and Planning, 119 (3), 289-305.

Krajewski, W. F., Lakshmi, V., Georgakakos, K. P. and Jain, S. C., 1991. A Monte-Carlo Study of Rainfall Sampling Effect on a Distributed Catchment Model. Water Resources Research, 27 (1), 119-128.

Krause, P., Boyle, D. P. and Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment, Advances in Geosciences, 5, 89-97.

Kristensen, K. J., Jensen, S. E., 1975. A model for estimating actual evapotranspiration from potential evapotranspiration. Nordic Hydrology, 6, 170-188.

Larson, L. W., 1971. Shielding Precipitation Gages from Adverse Wind Effects with Snow Fences. University of Wyoming Water Resources Research Institute, Water Resources Series, 25, 1-161.

Larson, L. W. and Peck, E. L., 1974. Accuracy of Precipitation Measurements for Hydrologic Modeling. Water Resources Research, 10 (4), 857-863.

Law, F., 1953. The estimation of the reliable yield of a catchment by correlation of rainfall and runoff. Journal Institute of Water Engineering, 7, 273-293.

Lee, Y. W., Chung, S. Y., Bogardi, I., Dahab, M. F. & Oh, S. E., 2001. Dose–response assessment by a fuzzy linear regression method. Water Science Technology, 43 (2), 133-140.

Lee, J., 1993. A Formal Approach to Hydrological Model Conceptualization. Hydrological Sciences Journal-Journal des Sciences Hydrologiques, 38 (5), 391-401.

Legates, D. and McCabe Jr, G., 1999. Evaluating the Use of "Goodness ‐ of ‐ Fit" Measures in Hydrologic and Hydroclimatic Model Validation, Water Resources Research., 35 (1), 233-241.

Lilly, A., Boorman, D. B. and Hollis, J. M., 1998. The development of a hydrological classification of UK soils and the inherent scale changes. Nutrient Cycling in Agroecosystems, 50, 299-302.

Liu, H-L., Chen, X., Bao, A-M., Wang, L., 2007. Investigation of groundwater response to overland flow and topography using a coupled MIKE SHE/MIKE 11 modeling

system for an arid watershed. Journal of Hydrology (Amsterdam), 347 (3-4), 448-459.

Loague, K. M. and Freeze, R. A., 1985. A comparison of rainfall-runoff techniques on small upland catchments. Water Resources Research, 21, 229-240.

LOCAR, 2000. Proposals for infrastructure and monitoring on the LOCAR Catchments, Appendices to the Report, NERC.

Lopes, V. L., 1996. On the effect of uncertainty in spatial distribution of rainfall on catchment modelling, Catena, 28 (1-2), 107–119.

Mackay, R., Cuthbert, M. O. Ash, H. and Tellam, J. H., 2006. Numerical upscaling to quantify aquifer recharge through glacial drift deposits, Shropshire, UK. In Bierkens, M. F. P., Gejrels, J. C. and KOvar, K. (eds) Calibration and reliability in groundwater modelling: from uncertainty to decision making, IAHS Publication 304, 52-57.

Madsen, H. and Kristensen, M., 2002. A multi-objective calibration framework for parameter estimation in the MIKE SHE integrated hydrological modelling system, in ModelCARE 2002. Proc. 4th Int. Conf. Calibration and Reliability in Groundwater Mod., Prague.

Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in Water Resources, 26 (2), 205-216.

Maier, H. R. and Dandy, G. C., 1997. Author's reply to comments by Fortin, V., Ouarda, T.B.M.J. and Bobee, B. on "The use of artificial neural networks for the prediction of water quality parameters" by Maier, H.R. and Dandy, G.C. Water Resources Research, 33 (10), 2425-2427.

Maréchal, D. and Holman, I., 2005. Development and application of a soil classification-based conceptual catchment-scale hydrological model, Journal of Hydrology, 312, 277-293.

McCuen, R. H., 1982. A Guide to Hydrologic Analysis Using SCS Methods, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

McCuen, R. H., Knight, Z. and Cutter, A. G., 2006. Evaluation of the Nash-Sutcliffe efficiency index. Journal of Hydrologic Engineering, 11 (6), 597-602.

McMichael, C. E., Hope, A. S. and Loaiciga, H. A., 2006. Distributed hydrological modelling in California semi-arid shrublands: MIKE SHE model calibration and uncertainty estimation. Journal of Hydrology, 317 (3-4), 307-324.

Mernild, S. H., Hasholt, B. and Liston, G., 2008. Climatic control on river discharge simulations, Zackenberg River Drainage basin, NE Greenland. Hydrological Processes, 22, 1932-1948.

Mertens, J., Madsen, H., Feyen, L., Jacques, D. and Feyen, J., 2004. Including prior information in the estimation of effective soil parameters in unsaturated zone modelling. Journal of Hydrology, 294 (4), 251-269.

Meteorological Office, 2008. Climate averages data, accessed August 2008, http://metoffice.gov.uk/ climate/uk/averages/

Michaud, J. and Sorooshian, S., 1994. Comparison of simple versus complex distributed runoff models on a mid-sized semiarid watershed. Water Resources Research, 21 (2), 593-605.

Minns, A. W. and Hall, M. J., 1996. Artificial neural networks as rainfall-runoff models. Hydrological Sciences Journal, 41, 399-417.

Molnar, D. K. and Julien, P. Y., 2000. Grid-Size Effects on Surface Runoff Modeling. Journal of Hydrologic Engineering, 5 (1), 8-16.

Monteith, J. L., 1965. Evaporation and environment. Symposium of the Society for Experimental Biology, Vol. XIX, 205-234.

Moore, R.V., 2007. The OpenMI Association Strategy Statement for the next decade, 2nd OpenMI Life Workshop and Associated Meetings, 20-21 November 2007, Wallingford, UK.

Moreda, F., Koren, V., Zhang, Z., Reed, S. and Smith, M., 2006. Parameterization of distributed hydrological models: learning from the experiences of lumped modeling. Journal of Hydrology, 320 (1-2), 218-237.

Murphy, J. M., Sexton, D. M. H., Jenkins, G. J., Boorman, P.M., Booth, B.B.B., Brown, C. C., Clark, R. T., Collins, M., Harris, G. R., Kendon, E. J., Betts, R. A., Brown, S. J., Howard, T. P., Humphrey, K. A., McCarthy, M. P., McDonald, R. E., Stephens, A., Wallace, C., Warren, R., Wilby, R., Wood, R. A., 2009. UK Climate Projections Science Report: Climate change projections. Met Office Hadley Centre, Exeter.

Nash, J. E. and Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part I - A discussion of principles, Journal of Hydrology, 10 (3), 282-290.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R. and Williams, J. R., 2005. Soil and Water Assessment Tool - Theoretical Documentation - Version 2005, Grassland, Soil and Water Research Laboratory, Agricultural Research Service and Blackland Research Center, Texas Agricultural Experiment Station, Temple, Texas.

Nelder, J. A. and Mead, R., 1965. A simplex method for function minimization, Computer Journal, 7, 308-313.

NERC 2008, The Lowland Catchment Research programme (LOCAR) http://catchments.nerc.ac.uk/

Ngo, L. L., Madsen, H. and Rosbjerg, D., 2007. Simulation and optimization modeling approach for operation of the HoaBinh reservoir, Vietnam. Journal of Hydrology, 336, 269-281.

NRFA, 2007. The National River Flow Archive; accessed April 2007. http://www.ceh.ac.uk/data/nrfa/index.html.

O'Connell, P. E., Beran, M. A., Gurney, R. J., Jones, D. A. and Moore, R. J., 1977. Methods for evaluating the U.K. raingauge network. Wallingford, Institute of Hydrology, 262. (IH Report No.40).

Omer, R. C., Nelson, E. J. and Zundel, A. K., 2003. Impact of varied data resolution on hydraulic modelling and floodplain delineation. Journal of the American Water Resources Association, 39 (2), 467-475.

Oudin, L. F., Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil and C. Loumagne., 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of Hydrology, 303, 290-306.

Page, T., Beven, K. J., Freer, J. and Neal, C., 2007. Modeling the chloride signal at Plynlimon, Wales, using a modified dynamic TOPMODEL incorporating conservative chemical mixing (with uncertainty), Hydrological Processes, 21, 292-307.

Perrin, C., Michel, C. and Andreassian, V., 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. Journal of Hydrology, 242 (3-4), 275-301.

Pietroniro, A., Leconte, R., Toth, B., Peters, D. L., Kouwen, N., Conley, F. M. and Prowse, T., 2006. Modelling climate change impacts in the Peace and Athabasca catchment and delta III – Integrated model assessment Hydrological Processes. 20 (19), 4231-4245, Special Northern Rivers Ecosystem Initiative Issue.

Pilling, C. and Jones, J. A. A., 1999. High resolution climate change scenarios: implications for British runoff. Hydrological Processes, 13 (17), 2877-2895.

Poole, D. K., Roberts, S. W. and Miller, P. C., 1981. Water utilization. In: Miller, P.C. (Ed.), Resource Use by Chaparral and Matorral: A Comparison of Vegetation Function in Two Mediterranean-type Ecosystems Ecological Studies No. 39. Springer, Berlin, 123-149.

Price, W. L., 1983. Global Optimization by Controlled Random Search, Journal of Optimization Theory and Applications, 40, 333-348.

Priestley, C. H. B. and Taylor, R. J., 1972. On the assessment of surface heat fluxes and evaporation using large-scale parameters. Monthly Weather Review, 100, 81–92.

Procter, C., Comber, L., Betson, M., Buckley, D., Frost, A., Lyons, H., Riding, A. and Voyce, K., 2006. Identifying crop vulnerability to groundwater abstraction: Modelling and expert knowledge in a GIS. Journal of Environmental Management, 81 (3), 296-306.

Prudhomme, C., Reynard, N. and Crooks, S., 2002. Downscaling of global climate models for flood frequency analysis: where are we now? Hydrological Processes, 16, 1137-1150.

QUEST, 2009. Quantifying Earth system processes and feedbacks for better informed assessments of alternative futures of the global environment, accessed November 2009. http://quest.bris.ac.uk/

Ragg, J. M., Beard, G. R., George, H., Heaven, F. W., Hollis, J. M., Jones, R. J. A., Palmer, R. C., Reeve, M. J., Robson, J. D. and Whitfield, W. A. D., 1984. Soils and their Use in Midland and Western England, Bulletin No. 12, Soil Survey of England and Wales, Harpenden.

Refsgaard, J. C., van der Sluijs, J. P., Brown, J. and van der Keur, P., 2006. A framework for dealing with uncertainty due to model structure error, Advance Water Resources, 29, 1586-1597.

Refsgaard, J. C., Storm, B. & Refsgaard, A. 1995. Recent developments of the Système Hydrologique Européen (SHE) towards the MIKE SHE. In: Modelling and Management of Sustainable Basin-Scale Water Resource Systems Boulder Symp., July 1995) (ed. by S. P. Simonovic, Z. Kundzewicz, D. Rosbjerg & K. Takeuchi), 425-434. IAHS Publ.no. 231.

Refsgaard, J.C. and Knudsen,J., 1996. Operational validation and intercomparison of different types of hydrological models. Water Resources Research, 32 (7): 2189-2202.

Refsgaard, J.C. and Storm, B., 1995. MIKE SHE, in Computer Models of Watershed Hydrology, Singh, V.P., Ed., Water Resources Publications, Colorado, USA, 809-846.

Refsgaard, J. C., 1997. Parameterisation, calibration and validation of distributed hydrological models. Journal of Hydrology, 198 (1-4), 69-97.

Refsgaard, J. C., 2000. Towards the formal approach to calibration and validation of models using spatial data. In: Grayson, R. and Blöschl, G., (eds.) Spatial patterns in catchment hydrology: observations and modelling. Chapter 13, 329.

Robinson, A. C. and Rodda, J. C., 1969. Rain, wind and the aerodynamic characteristics of raingauges. Meteorology Magazine, 98, 113-120.

Rubarenzya, M. H., Graham, D., Feyen. J., Willems, P. and Berlamont, J., 2007. A site-specific land and water management model in MIKE SHE. Nordic Hydrology, 38 (4-5), 333-350.

Ruosteenoja, K., Carter, T. R., Jylhä, K. and Tuomenvirta, H., 2003. Future climate in world regions: an intercomparison of model-based projections for the new IPCC emissions scenarios, The Finnish Environment, Finnish Environment Institute, 83.

Russo, D. and Bresler, E., 1980. Scaling soil hydraulic properties of a heterogenous soil. Soil Science Society of America Journal 44, 681-684.

Sahoo, G. B., Ray, C. and De Carlo, E. H., 2006. Calibration and validation of a physically distributed hydrological model, MIKE SHE, to predict streamflow at high frequency in a flashy mountainous Hawaii stream. Journal of Hydrology, 327 (1-2), 94-109.

Saltelli, A., 2000. Fortune and future of sensitivity analysis. In: A. Saltelli, K. Chan and E.M. Scott, Editors, Sensitivity Analysis, Wiley, Chichester, 421-426.

Santhi, C., Srinivasan, R., Arnold, J. G. and Williams, J. R., 2006. A modeling approach to evaluate the impacts of water quality management plans implemented in a watershed in Texas, Environmental Modelling and Software, 21, 1141-1157.

Saulnier, G. M., Obled, C. and Beven, K., 1997. Analytical compensation between DTM grid resolution and effective values of saturated hydraulic conductivity within the TOPMODEL framework. Hydrological Processes, 11 (9), 1331-1346.

Schenk, H. J. and Jackson, R. B., 2002. Rooting depths, lateral root spreads, and below-ground/above-ground allometries of plants in water-limited ecosystems. Journal of Ecology, 90, 480-494.

Schlesinger, S., Crosbic, R. E., Gagné, R. E., Innis, G. S., Lalwani, C. S., Loch, J., Sylvester, J., Wright, R. D., Kheir, N. and Bartos, D., 1979. Terminology for model credibility. Simulation, 32 (3), 103-104.

Sefton, C. E. M. and Boorman, D. B., 1997. A regional investigation of climate change impacts on UK streamflows, Journal of Hydrology, 195, 26-44.

Shah, S. M. S., O'Connell, P. E. and Hosking, J. R. M., 1996. Modeling the effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. Journal of Hydrology, 175, 89-111.

Shamseldin, A.Y., 1997. Application of a neural network technique to rainfall-runoff modeling. Journal of Hydrology (Amsterdam), 199 (3/4), 272-294.

Shaw, E. M., 1994. Hydrology in Practice (Third Edition), Chapman and Hall, London.

Shepley, M. G., Streetly, M., Voyce, K. and Bamford, F., 2009. Management of stream compensation for a large conjunctive use scheme, Shropshire, United Kingdom. Water and Environment Journal, 23, 263-271.

Sherman, L. K., 1932. Streamflow from rainfall by the unit hydrograph method, Eng. News Record, 108, 501-505.

Shrestha, R., Tachikawa, Y. and Takara, K., 2006. Input data resolution analysis for distributed hydrological modeling. Journal of Hydrology, 319 (1-4), 36-50.

Sieber, A. and Uhlenbrook, S., 2005. Sensitivity analyses of a distributed catchment model to verify the model structure. Journal of Hydrology, 310 (1-4), 216-235.

Silberstein, R. P., 2006. Hydrological models are so good, do we still need data? Environmental Modelling & Software, 21 (9), 1340-1352.

Singh, C. R., Thompson, J. R., French, J. R., Kingston, D. G., and Mackay, A. W., 2010. Modelling the impact of prescribed global warming on water resources of headwater catchments of the Irrawaddy River and their implications for Loktak Lake, northeast India. Hydrology Earth Systems Science Discussions, 7, 2781-2828.

Singh, V. P. (Editor), 1995. Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado.

Sivapalan, M. and Kalma, J. D., 1995. Scale problems in hydrology: the contributions of the Robertson workshop. In: Kalma, J. D., Sivapalan, M. (Eds.), Scale Issues in Hydrological Modeling. Wiley, England, 1-8.

Smerdon, B. D., Allen, D. M., Grasby, S. E. and Berg, M. A., 2009. An approach for predicting groundwater recharge in mountainous watersheds. Journal of Hydrology, 365,156-172.

Sonnenborg, T. O., Christensen, B. S. B., Nyegaard, P., Henriksen, H. J. and Refsgaard, J. C., 2003. Transient modeling of regional groundwater flow using parameter estimates from steady-state automatic calibration. Journal of Hydrology, 273 (1-4), 188-204.

Spear, R. C. and Hornberger, G. M., 1980. Eutrophication in Peel Inlet-II Identification of critical uncertainties via generalized sensitivity analysis, Water Resources, 14, 43-99.

Stadnyk, T. A., St Amour, N. A., Kouwen, N., Edwards, T. W. D., Pietroniro, A. and Gibson, J. J., 2005. A groundwater seperation study in boreal wetland terrain: the WATFLOOD hydrological model compared with stable isotope tracers. Isotopes in Environmental and Health Studies (in press).

Steele-Dunne, S., Lynch, P., McGrath, R., Semmler, T., Wang. S., Hanafin. J. and Nolan, P., 2008. The impacts of climate change on hydrology in Ireland. Journal of Hydrology, 356, 28-45.

Stephenson, P. M., 1968. Objective assessment of adequate numbers of rain gauges for estimating areal rainfall depths. Proceedings, IASH General Assembly (Bern 1967), Publ. No. 78, Bern, 252-264.

Stephenson, G. R. and Freeze, R. A., 1974. Mathematical Simulation of Subsurface Flow Contributions to Snowmelt Runoff, Reynolds Creek Watershed, Idaho. Water Resources Research, 10 (2), 284-294.

Stisen, S., Jensen, K. H., Sandholt, I. and Grimes, D. I. F., 2008. A remote sensing driven distributed hydrological model of the Senegal River basin. Journal of Hydrology, 354, 131-148.

Storm, B., Jørgensen, G. H. and Styczen, M., 1987. Simulation of water flow and soil erosion processes with a distributed physically-based modelling system, in Forest Hydrology and Watershed Management Proceedings of the Vancouver Symposium, August 1987, IAHS-AISH Publ. No. 167.

Su, M., Stolte, W. J. and van der Kamp, G., 1997. Modelling wetland hydrology using SLURP. Proc. Scientific Meeting of the Canadian Geophysical Union, Banff, Alberta, 198.

Sumner, G., 1988. Precipitation: Process and Analysis. J. Wiley & Sons, Chichester, UK.

Sun, X., Mein, R. G., Keenan, T. D. and Elliott, J. F., 2000. Flood estimation using radar and raingauge data. Journal of Hydrology, 239, 4-18.

Thiessen, A. H., 1911. Precipitation for large areas, Mon. Weather Review, 39, 1082-1084.

Thodsen, H. 2007. The influence of climate change on stream flow in Danish rivers. Journal of Hydrology, 333, 226-238.

Thompson, J. R., Gavin, H., Refsgaard, A., Refstrup Sørenson, H. and Gowing, D. J., 2009. Modelling the hydrological impacts of climate change on UK lowland wet grassland. Wetlands Ecology and Management, 17 (5), 503-523.

Thompson, J.R. and Hollis, G.E. 1997. Wetlands and Integrated River Basin Management: Experiences in Asia and the Pacific: Part I, Wetlands International: Asia-Pacific, Kuala Lumpur, Malaysia and United Nations Environment Programme, Nairobi, Kenya.

Thompson, J. R., Sorenson, H. R., Gavin, H. and Refsgaard, A., 2004. Application of the coupled MIKE SHE/MIKE 11 modelling system to a lowland wet grassland in southeast England. Journal of Hydrology, 293 (1-4), 151-179.

Thornthwaite, C.W., 1948. An approach towards a rational classification of climate. Geographical Review, 38, 55–94.

Thorsen, M., Refsgaard, J. C., Hansen, S., Pebesma, E., Jensen, J. B. and Kleeschulte, S., 2001. Assessment of uncertainty in simulation of nitrate leaching to aquifers at catchment scale. Journal of Hydrology, 242 (3-4), 210-227.

Toth, B., Pietroniro, A., Conly, F. M. and Kouwen, N., 2006. Modelling climate change impacts in the Peace and Athabasca catchment and delta: I–hydrological model application. Hydrological Processes, 20 (19), 4197-4214.

Tsang, C. F., 1991. The Modeling Process and Model Validation. Ground Water, 29 (6), 825-831.

UNESCO, 2003. Water for people, water for life: The United Nations world water development report, UNESCO Publishing, Paris / Berghahn Books, Oxford.

United Nations, 1992. Chapter 18 of Agenda 21 Report of UN, Rio de Janeiro, Report of the UN Conference on Environment and Development, Impact of Climate Change on Water Resources.

Van der Linden, S. and Woo, M. K., 2003a. Application of hydrological models with increasing complexity to subarctic catchments. Journal of Hydrology, 270 (1-2), 145-157.

Van der Linden, S. and Woo, M. K., 2003b. Transferability of hydrological model parameters between basins in data-sparse areas, subarctic Canada. Journal of Hydrology, 270 (3-4), 182-194.

Vazquez, R. F, Willems, P. and Feyen, J., 2008. Improving the predictions of a MIKE SHE catchment-scale application by using a multi-criteria approach. Hydrology Process, 22 (13), 2159-2179.

Vazquez, R. F. and Feyen, J., 2003. Effect of potential evapotranspiration estimates on effective parameters and performance of the MIKE SHE-code applied to a medium-size catchment. Journal of Hydrology, 270 (3-4), 309-327.

Vazquez, R. F. and Feyen, J., 2007. Assessment of the effects of DEM gridding on the predictions of basin runoff using MIKE SHE and a modelling resolution of 600 m. Journal of Hydrology, 334 (1-2), 73-87.

Vazquez, R. F., Feyen, L., Feyen, J. and Refsgaard, J. C., 2002. Effect of grid size on effective parameters and model performance of the MIKE-SHE code. Hydrological Processes, 16 (2), 355-372.

Vogel, R. M. and Sankarasubramanian, A., 2003. Validation of a watershed model without calibration. Water Resources Research, 39 (10), 1292.

Ward, R. C. and Robinson, M., 1990. Principles of Hydrology, (Third Edition), McGraw-Hill Book Company, London.

Wiesner, C. J., 1970. Hydrometeorology. Chapman and Hall, London, UK.

Wilby, R. L., 2005. Uncertainty in water resource model parameters used for climate change impact assessment. Hydrological Processes, 19 (16), 3201-3219.

WMO., 1994. Guide to Hydrological Practices, Data Acquisition and Processing, Analysis, forecasting and other applications. WMO-No. 168, Fifth Edition.

Wolock, D. M. and Hornberger, G. M., 1991. Hydrological effects of changes in atmospheric carbon dioxide levels. Journal of Forecasting, 10, 105-116.

Woo, M-K. and Thorne, R., 2006. Snowmelt contribution to discharge from a large mountainous catchment in subarctic Canada. Hydrological Processes, 20, 2129-2139.

Wood, E. F., Sivapalan, M., Beven, K. and Band, L., 1988. Effects of Spatial Variability and Scale with Implications to Hydrologic Modeling. Journal of Hydrology, 102 (1-4), 29-47.

Woolhiser, D. A., Smith, R. E. and Goodrich, D. C., 1990. A kinematic runoff and erosion manual: documentation and user manual. US Department of Agriculture, Washington, DC, ARS 77, 130.

Xavier, L. N. R., De-Magalhaes, A. A. A. and Filho, O. C. R., 2006. Bayesian kriging and GLUE applied to estimation of uncertainty due to precipitation representation in hydrological modelling, IAHS-AISH-Publication, 303, 90-98.

Xevi, E., Christiaens, K., Espino, A., Sewnanden, W., Mallants, D., Sorensen H. and Feyen, J., 1997. Calibration, validation and sensitivity analysis of the MIKE-SHE model using the Neuenkirchen catchment as case study. Water Resources Management, 11, 219-242.

Yang, D., Herath, S. and Musiake, K., 2000. Comparison of different distributed hydrological models for characterization of catchment spatial variability. Hydrological Processes, 14 (3), 403-416.

Zhang, W. H. and Montgomery, D. R., 1994. Digital Elevation Model Grid Size, Landscape Representation, and Hydrologic Simulations. Water Resources Research, 30 (4), 1019-1028.

Zhang, Z., Wang, S., Sun, G., McNulty, S. G., Zhang, H., Li, J., Zhang, M., Klaghofer, E. and Strauss, P., 2008. Evaluation of the MIKE SHE Model for Application in the Loess Plateau, China. Journal of the American Water Resources Association, 44 (5), 1108-1120.